



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

基于神经网络的光谱成像与重建技术研究

作者姓名：王梓

指导教师：王建宇 研究员

中国科学院上海技术物理研究所

学位类别：工学博士

学科专业：物理电子学

培养单位：中国科学院上海技术物理研究所

2025 年 6 月

Computational Spectral Imaging and Reconstruction Based on
Neural Networks

A dissertation submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor of Engineering
in Physical Electronics

By

Zi Wang

Supervisor: Professor Jianyu Wang

Shanghai Institute of Technical Physics, Chinese Academy of Sciences

June, 2025

中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。承诺除文中已经注明引用的内容外，本论文不包含任何其他个人或集体享有著作权的研究成果，未在以往任何学位申请中全部或部分提交。对本论文所涉及的研究工作做出贡献的其他个人或集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期：

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关收集、保存和使用学位论文的规定，即中国科学院大学有权按照学术研究公开原则和保护知识产权的原则，保留并向国家指定或中国科学院指定机构送交学位论文的电子版和印刷版文件，且电子版与印刷版内容应完全相同，允许该论文被检索、查阅和借阅，公布本学位论文的全部或部分内容，可以采用扫描、影印、缩印等复制手段以及其他法律许可的方式保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

光谱成像技术通过获取目标场景的三维光谱数据立方体,为物质识别、环境监测、医学诊断及军事遥感等领域提供了有力的表征手段。然而,传统光谱成像依赖分光设备和机械扫描,存在成像速度慢、设备体积大及难以处理散射场景等局限。近年来,计算光谱成像技术通过光学编码和算法解码的协同设计,显著提升了成像性能,推动了系统的小型化。尽管如此,当前的方法仍然面临硬件依赖性大、重建质量有限以及数据驱动算法训练样本稀疏等挑战。神经网络在特征提取方面的优势,使其在优化光谱成像全流程中成为研究重点。

因此,本文将围绕计算光谱成像系统的编码-解码框架,基于神经网络的方法,研究了神经网络在计算光谱重建问题中的若干难点,包括训练数据不足、编码设计受制和全局光谱成像的重建问题,分别从有监督学习、混合监督学习以及预训练自监督学习三条技术路线出发,创新性地、递进地提出三种解决方案:

(1)提出了结合物理仿真模型和有监督学习,解决缺乏训练数据的问题,开发了基于像素的分层光谱变换网络(Hierarchical Spectral Transformer, HST)与 Cross-Talk Transformer,在长波红外多孔径快照光谱成像及工业气体泄漏监测中取得了验证效果,证明有效提升了低质量光谱图像的光谱分辨率与噪声鲁棒性,并实现高精度的气体识别和浓度监测。相关仪器已经投入实际工业园区监测的应用中,达到国内领先水平。

(2)提出了基于神经光谱场的混合监督学习策略,采用设计受限的超小型化电压可调谐光谱探测器验证了方法的有效性,实现了 0.30nm 波长精度和 10nm 的光谱分辨率,显示出优秀的光谱成像能力,并且与传统基于优化的重建方法对比体现出了对噪声的鲁棒性,证明所提出的方法达到国际领先水平。

(3)提出了基于预训练和光谱相位场的自监督学习光谱重建,不再需要额外的训练数据集,同时融合了相位场去噪和预训练光谱解码能力,在非视域条件上进行实验论证,在国际上首次实现高质量的非视域场景下的光谱图像重建方法和系统,达到了 10nm 的光谱分辨率和 3cm@1m 的空间分辨率,具备非视域场景下的物质识别和材质识别能力,达到国际领先水平。

本文系统探讨了神经网络在光谱成像领域的应用,对计算光谱成像编码—解码框架进行了深入研究,并提出了三种基于仿真驱动有监督、物理约束混合监督及自监督优化的重建方法,为不同场景下的光谱成像提供了可扩展解决方案,并在气体监测、微型光谱分析及非视距探测等方向实现技术验证。未来研究可向多模态融合、轻量化网络设计及物理可解释模型构建延伸,以进一步提升光谱成像系统的智能化及实用性。

关键词: 计算光谱成像; 神经网络; 有监督学习; 自监督学习; 预训练; 混合监督学习

Abstract

Spectral imaging technology, by capturing three-dimensional spectral data cubes of target scenes, has become an invaluable tool in fields such as materials identification, environmental monitoring, medical diagnostics, and military remote sensing. However, traditional spectral imaging—relying on dispersive optical devices and mechanical scanning—suffers from slow imaging speeds, bulky apparatus, and challenges in handling scattering scenes. In recent years, computational spectral imaging, through the synergistic design of optical encoding and algorithmic decoding, has markedly enhanced imaging performance and fostered system miniaturization. Nonetheless, current methodologies continue to grapple with significant hardware dependency, limited reconstruction quality, and a dearth of training samples for data-driven algorithms. Leveraging the prowess of neural networks in feature extraction, these have emerged as a focal point in optimizing the entire spectral imaging process.

Accordingly, this paper centers on the encoding-decoding framework of computational spectral imaging systems, investigating several challenges inherent in neural network-based spectral reconstruction. These include insufficient training data, constraints in encoding design, and the overarching issue of global spectral image reconstruction. We systematically propose three innovative and progressively refined solutions developed along three technical avenues: supervised learning, hybrid supervised learning, and pre-trained self-supervised learning.

(1) We introduce a method that amalgamates physical simulation models with supervised learning to overcome the scarcity of training data. By developing a pixel-based Hierarchical Spectral Transformer(HST) integrated with a Cross-Talk Transformer, we have validated the technique in long-wave infrared multi-aperture snapshot spectral imaging and industrial gas leakage monitoring. The approach significantly enhances the spectral resolution and noise robustness of low-quality spectral images, thereby achieving highly accurate gas identification and concentration monitoring. The related instruments have already been deployed in industrial park monitoring, attaining a leading domestic standard.

(2) We propose a hybrid supervised learning strategy based on neural spectral fields. Utilizing a design-constrained, ultra-miniaturized voltage-tunable spectral detector, we have demonstrated its efficacy in achieving 0.30 nm wavelength accuracy and 10 nm spectral resolution. The method exhibits exceptional spectral imaging capabilities, and its robustness to noise is clearly evidenced when compared to traditional optimization-based reconstruction methods, establishing the approach as international state-of-the-art.

(3) We advocate a self-supervised learning spectral reconstruction framework grounded in pre-training and spectral phase field concepts, eliminating the need for additional training datasets. By integrating phase field denoising with pre-trained spectral decoding capabilities, we have experimentally validated this approach under non-line-of-sight conditions. This marks the first international achievement in high-quality spectral image reconstruction in non-line-of-sight scenarios, reaching a spectral resolution of 10 nm and a spatial resolution of 3 cm at 1 m. The system is capable of material and substance recognition in non-line-of-sight scenarios, reaching leading international standards.

This paper systematically examines the application of neural networks in the realm of spectral imaging, offering an in-depth exploration of the computational spectral imaging encoding-decoding framework. It proposes three reconstruction methods based on simulation-driven supervised learning, physics-constrained mixed supervision, and self-supervised optimization, thereby providing scalable solutions for spectral imaging across diverse scenarios. These advancements have been technically validated in areas such as gas monitoring, micro-spectroscopy, and non-line-of-sight detection. Future research may extend toward multimodal fusion, lightweight network design, and the development of physically interpretable models to further enhance the intelligence and practicality of spectral imaging systems.

Key Words: Computational spectral imaging; Neural networks; Longwave-infrared imaging; Miniaturized spectrometer; Non-line-of-sight imaging; Hybrid-supervised imaging

目 录

第 1 章 绪论	1
1.1 研究背景和意义	1
1.2 相关领域研究现状	3
1.2.1 光谱重建与优化方法	3
1.2.2 光谱重建与神经先验方法	4
1.2.3 光谱重建与神经网络方法	6
1.3 光谱重建存在的问题与挑战	9
1.4 本文主要研究内容	12
1.5 本文章节安排	13
第 2 章 基于物理仿真模型的有监督学习光谱重建	15
2.1 引言	15
2.2 快照式光谱成像系统	15
2.2.1 多孔径快照式光谱成像	15
2.2.2 长波红外多孔径快照式光谱成像系统	16
2.3 像素级长波红外光谱图像重建	19
2.3.1 信号传输与噪声模型	19
2.3.2 网络结构	19
2.3.3 位置编码	20
2.3.4 分层表示	20
2.3.5 注意力机制	21
2.3.6 实现细节	21
2.4 高鲁棒性气体种类和浓度监测	22
2.4.1 气体辐射传输模型	22
2.4.2 神经网络结构	25
2.4.3 实现细节	27
2.4.4 神经网络实现	27
2.5 图像重建实验结果	28
2.5.1 评估指标	28
2.5.2 仿真数据结果	29
2.5.3 实验室测试数据结果	30
2.5.4 外场测试数据结果	34
2.6 气体监测实验结果	36

2.6.1 实验装置	36
2.6.2 实验数据	37
2.6.3 评价指标	38
2.6.4 仿真数据结果	39
2.6.5 实验数据结果	39
2.7 本章小结	44
第 3 章 基于神经光谱场的混合监督学习光谱重建	47
3.1 引言	47
3.2 计算微型光谱仪	47
3.2.1 计算微型光谱仪的设计	47
3.2.2 微型光谱仪的工作机制	49
3.2.3 微型光谱仪的电学性能	50
3.3 光谱测量流程	52
3.4 光谱重建	56
3.4.1 重建问题描述	56
3.4.2 光谱重建流程	57
3.4.3 响应函数表征	59
3.4.4 神经网络结构	59
3.4.5 合成数据集构建	60
3.4.6 神经网络训练	61
3.5 重建结果	62
3.5.1 峰谱重建结果	62
3.5.2 光谱成像结果	65
3.5.3 测量误差的影响	67
3.5.4 宽带重建	67
3.5.5 多峰重建	73
3.6 本章小结	77
第 4 章 基于神经瞬态场的的自监督学习多光谱重建	79
4.1 引言	79
4.2 系统分析	80
4.2.1 被动式与主动式	80
4.2.2 共焦与非共焦	81
4.3 非视距成像的传统方法	82
4.3.1 多光谱神经瞬态场	83
4.3.2 共焦与非共焦实施	84

4.3.3 共焦系统	85
4.3.4 非共焦系统	86
4.4 噪声模型和损失函数	86
4.4.1 单光子探测器噪声模型	86
4.4.2 基于模型的预测和泊松损失	87
4.4.3 泊松损失	88
4.5 实验结果	89
4.5.1 合成数据生成	89
4.5.2 实际数据测量	90
4.5.3 单光子模型的验证	90
4.5.4 定性比较	91
4.5.5 定量比较	95
4.6 结论	96
第 5 章 基于预训练和光谱相位场的自监督学习光谱重建	99
5.1 引言	99
5.1.1 非视距光谱成像	99
5.1.2 非视距光谱成像系统	100
5.2 前向模型	101
5.2.1 扫描模式设计	102
5.3 神经光谱相位场	103
5.3.1 瑞利-索末菲衍射相位场	105
5.3.2 神经场中的光谱解码器	105
5.4 实验设置	106
5.5 成像系统	107
5.5.1 光谱和空间定标	107
5.5.2 激光功率泄漏	108
5.6 实验结果	109
5.6.1 仿真数据结果	109
5.6.2 实测数据结果	119
5.6.3 结论	124
第 6 章 结论与展望	127
参考文献	129
致谢	139
作者简历及攻读学位期间发表的学术论文与其他相关学术成果	141

图目录

图 1-1	论文研究内容框架	11
图 2-1	多孔径快照式光谱成像系统原理	17
图 2-2	非制冷快照红外光谱仪设计与成像效果	18
图 2-3	像素级分层光谱 Transformer 结构示意图	20
图 2-4	Class-talk Transformer 结构示意图	25
图 2-5	Transformer 特征提取器的结构示意图	26
图 2-6	训练、验证和测试过程中的损失曲线	28
图 2-7	HST 重建四种不同类型的光谱定性分析	31
图 2-8	使用仿真数据集比较不同方法	32
图 2-9	通过实验室数据比较四种不同气体传输光谱的不同方法	34
图 2-10	外场实验展示 HST 的应用。	36
图 2-11	非制冷快照红外光谱仪气体实验装置	37
图 2-12	Class-talk Transformer 在仿真数据集上的气体种类识别结果	39
图 2-13	Cross-talk Transformer 在模拟数据集上的浓度预测结果	40
图 2-14	Class-talk Transformer 在实测数据集上的气体种类识别结果	42
图 2-15	Cross-talk Transformer 在实验数据集上的浓度预测结果。	43
图 2-16	Cross-talk Transformer、CNN 和 MLP 在浓度预测任务上的比较。	44
图 3-1	渐变带隙 PN 结光谱仪的示意图和电性能。	49
图 3-2	渐变带隙 PN 结光谱仪的示意图和电性能。	51
图 3-3	渐变带隙 PN 结光谱仪的显微镜图像	52
图 3-4	渐变带隙 PN 结光谱仪的光谱测量	53
图 3-5	渐变带隙 PN 结光谱仪的光电流	54
图 3-6	神经光谱场	54
图 3-7	提取特征的低维分析	55
图 3-8	重建光谱	56
图 3-9	渐变带隙 PN 结光谱仪的光谱重建流程	58
图 3-10	单峰光谱重建结果	63
图 3-11	LED 光谱重建结果	63
图 3-12	双峰光谱重建结果	64
图 3-13	高精度光谱重建结果	65
图 3-14	光谱重建装置	66

图 3-15	光谱重建流程	66
图 3-16	光谱重建结果	67
图 3-17	评估不同方法在测试阶段对噪声的鲁棒性	68
图 3-18	重建性能与峰宽之间关系的评估	69
图 3-19	重建性能与峰宽之间关系的评估	71
图 3-20	重建性能与峰宽之间关系的评估	72
图 3-21	重建性能与峰宽之间关系的评估	73
图 3-22	多峰光谱重建	75
图 3-23	多峰光谱重建	76
图 4-1	多光谱非视距成像	80
图 4-2	提出的 MSNeTF 流程	82
图 4-3	共焦与非共焦的系统设置	85
图 4-4	带环境光建模的 MSNeTF 结果	87
图 4-5	泊松损失函数的结果	88
图 4-6	硬件系统	91
图 4-7	共焦和非共焦场景重建结果	92
图 4-8	人体模型上的结果	93
图 4-9	双色纸板上的结果	93
图 4-10	零食包装袋上的结果	94
图 4-11	皮卡丘上的结果	94
图 5-1	三种不同的光谱非视距成像框架	101
图 5-2	扫描模式设计	103
图 5-3	神经光谱相位场的方法流程	104
图 5-4	非视距光谱成像系统	107
图 5-5	非视距光谱成像系统光谱定标	108
图 5-6	超连续激光系统中的功率泄漏	109
图 5-7	窄带反射率物体的重建	110
图 5-8	窄带反射率物体的完整光谱数据重建 (550nm-690nm)	111
图 5-9	窄带反射率物体的完整光谱数据重建 (400nm-540nm)	112
图 5-10	宽带反射率物体的重建	114
图 5-11	窄带反射贴片的重建	115
图 5-12	窄带反射贴片的重建	115
图 5-13	宽带'NLOS' 字母重建	116
图 5-14	方法在不同的数据分辨率和噪声下的重建	117
图 5-15	消融实验	118
图 5-16	不同材质物体的重建	118

图 5-17	实测反光材质'NLOS' 字母的重建	120
图 5-18	实测反光材质叶片的重建	121
图 5-19	实测真假荷叶的重建	122
图 5-20	实测荷叶的完整光谱数据重建	123
图 5-21	实测马的重建	124

表目录

表 2-1	在不同噪声水平下使用仿真数据集比较不同方法的表现。	33
表 2-2	使用实验室数据定量比较不同方法。	35
表 2-3	Class-talk Transformer 在模拟和实验数据集上的识别性能。	41
表 2-4	气体识别和浓度预测方法准确率比较。	43
表 4-1	贴片场景定量重建结果。	95
表 4-2	兔子场景的定量重建结果。	95
表 4-3	人体场景的定量重建结果。	96
表 4-4	不同数据的设定与参数	96

符号列表

字符

Symbol	Description	Unit
y	the measurement	1
Φ	the sensing matrix	1
f	the hyperspectral imaging vector	1
n	the detector noise	1
Ψ	the inverse sparse transform	1
α	the regularization parameter	1
R	the regularization term	1
L_t	the radiance at the lens	$\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1}$
L	the radiance at the detector	$\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1}$
τ	the spectral response	1
L_p	the background radiation	$\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1}$
ϕ	the quantum efficiency	1
N_d	the dark digital number	1
N_q	the quantification efficiency	1
$\mathcal{N}$	the Gaussian noise random variable	1
I	the measured intensity	1
m	the feature position	1
M	the maximum encoding dimension	1
S	the concatenated input	1
Q	the query	1
K	the key	1
V	the value	1
$\hat{\alpha}$	the deep corelation	1
B	the attention	1

算子

Symbol	Description
$\ \cdot\ _{TV}$	the total variation
$\ \cdot\ _*$	the nuclear norm

$\ \cdot\ _1$	the L1 norm
$\mathcal{P}$	the poisson noise operater

缩写

FPA	Focal Plane Array
CASSI	Coded Aperture Snapshot Spectral Imager
AOTF	Acousto-optic Tunable Filter
LCTF	Liquid Crystal Tunable Filters
MEMS	Micro-electro-mechanical System
FWHM	Full Width at Half Maximum
LOS	Line-of-sight
NLOS	Non-line-of-sight
3D	Three Dimension
TOF	Time of Flight
SPAD	Single Photon Avalanche Diode
TCSPC	Time-correlated single-photon counting
SNR	Signal-to-Noise Ratio
BP	Back-projection
FBP	Filtered Back-projection
LCT	Light-cone Transform
PF	Phasor Fields
SOCR	Signal-Object Collaborative Regularization
TV	Total Variation
RIP	Restricted Isometry Property
RBF	Radial Basis Function
HST	Hierarchical Spectral Transformer
LWIR	Long-wave Infrared
USIRS	Uncooled Snapshot Infrared Spectrometer
CNN	Convolutional Neural Networks
MLP	Multi-Layer Perceptron
FCN	Fully Connected Network
RMSE	Root Mean Square Error
PSNR	Peak signal-to-noise ratio
MAE	Mean Absolute Error

NSF	Neural Spectral Fields
LED	Light-emitting Diode
NeTF	Neural Transient Fields
NeRF	Neural Radiance Field
PS	Photometric stereo
PDF	Probability Density Function
MCMC	Markov Chain Monte Carlo
MSFA	Multi-spectral Filter Array
SVD	Singular Value Decomposition
SSIM	Structure Similarity Index Measure

第1章 绪论

1.1 研究背景和意义

物质在不同波长下具有不同的光学表现,如吸收和反射。通过记录目标的在多个波长下的光学强度数据,称之为物质的光谱。大多数物质都具有独特的光谱特征,因此光谱也被称为物质的指纹^[1,2]。通过探测物质的光谱能够确定物质的种类、组分和含量,这使得光谱成为物质最强大的表征工具之一。获取光谱信息最常见的方式是光谱仪。光谱仪通过分光获取某个像素点的光谱。而对于空间上有差异性的目标,通常希望获取目标在某个区域内每一个像素点的光谱信息,因此孵化了光谱成像技术。

光谱成像技术结合了光谱仪和成像技术,能够获取目标场景的三维数据立方体。从光谱维上来看,能够获取图像每个像素点光谱信息;而从空间维上来看,光谱成像是针对同一个场景在不同波长成像,每个波段都能得到一幅图像,这些图像的集合就构成了光谱图像。光谱成像技术不仅能获取空间信息,还能获取光谱信息,因此能够对目标区域的每一个像素点进行分析其物质的化学成分、物理状态等详细信息,这在许多领域都有重要的应用,如在环境监测时分析污染气体排放的种类、浓度和范围^[3],在医学诊断中分析病灶的种类、严重程度和区域^[4],在军事侦查中将关键目标与复杂背景进行区分^[5],在遥感监测对每一个像素的地物进行分类等^[6]。

光谱成像的核心是三维光谱数据立方体的获取,其核心矛盾在于如何使用二维的焦平面阵列 (Focal Plane Array, FPA) 获取到三维的光谱数据立方体。早期一系列的工作通过对三维的光谱数据立方体进行多次二维的测量来获取光谱图像。如推扫式高光谱成像仪采取了“用时间换空间的思想”,使用狭缝限制场景的入射光,再通过色散器件进行分光,单次曝光用二维的焦平面阵列获取目标场景一条线段的光谱信息,也是二维的空间-光谱信息,然后通过控制狭缝在场景的不同线段上进行扫描,最终将所有线段上的空间-光谱数据叠加起来,形成三维的光谱数据立方体^[7]。而滤光片式光谱成像仪则是采取了“用时间换光谱”的思想,每次在成像系统内使用不同的滤光片,通过多次采集目标的单波段图像,并将数据叠加形成三维光谱数据立方体^[8]。

然而这些早期的光谱成像技术却存在一些局限性。第一,传统光谱成像通过牺牲了时间进行多次测量解决二维探测器和三维光谱数据立方体之间的矛盾,因此成像速度慢,成像帧率低,难以对运动状态的目标进行成像^[9]。第二,光谱成像仪器通常都需要棱镜、光栅等色散器件实现分光,并且需要机械结构实现扫描或者切换滤光片,因此具有较为复杂的光学和机械结构,导致了较大的体积和重量,难以小型化和商业化^[10]。第三,光谱成像技术通过收集直接从目标反射或者发出的光,光学上受限于弹道光成像,因此对于复杂场景中存在的非弹道光,如光学散射或者漫反射,则难以处理^[11,12]。

近年来的计算光谱成像技术打破了对“直接测量”和“所见即所得”的传统光谱成像技术的路径依赖^[13]。计算光谱成像的最大特征是探测器获取到的图像可能视觉看起来与最终需要的光谱图像非常不相似，这是因为在图像获取的光学前端上引入了光学编码的概念，同时在后端处理时再通过计算机技术解码出光谱图像。计算光谱成像技术将图像获取的光学技术和图像处理的计算机技术结合起来，从统一的视角来看待图像获取和图像处理。使用计算光谱成像，人们设计出的仪器不局限于直接测量来得到图像，并且将强大的算法引入到成像系统中，其能力远远超过传统光学直接获取的物理限制。同样，计算成像使得能设计出与传统方法相比，以较小的大小、重量、功耗或成本提供所需功能的仪器，或者是能够实现超出传统方法想象的特殊成像能力。

计算光谱成像的框架可以粗略地分为两个部分：编码和解码。编码可定义为将三维光谱数据立方体在某个光学系统的功能下投影为焦平面阵列能够完全接收的二维数据。不同的光学系统所实现的编码形式很可能完全不同，如编码孔径快照式光谱成像（Coded Aperture Snapshot Spectral imaging, CASSI）系统是通过图像空间编码后再进行一维色散分光，最后在光谱维度上压缩叠加^[14,15]。从这个视角看传统光谱成像系统也可被视为特定条件下的计算光谱成像系统：编码函数为狄拉克函数，因此解码过程极为简单。并且由于编码过程总是需要通过物理的光学系统来实现，因此设计上受到较大的约束——最优的编码方式在理论上是存在的，却难以在光学上实现。因此，计算光谱成像的编码设计呈现出百花齐放、万物生机的多样性，针对不同场景具有各种特定设计的编码过程。如对动态目标进行成像、以实时性为主要目标的计算光谱成像系统，可以通过多孔径分光实现与传统单波段成像一致的成像帧率^[16]。而对于以小型化为主要目标的系统，可以通过外加电压控制实现探测器光谱响应的改变，从而对光谱进行编码^[17]。亦或者是对于非视距场景下的计算光谱成像，则是需要分析非视距的物理过程，建立探测器和目标之间的信号模型，才能了解其编码形式^[18,19]。因此，计算光谱成像的编码不存在通用的唯一解，编码形式是高度多样化的。

通过编码后探测器得到的二维数据需要再通过解码成为人能够理解的三维高光谱图像，解码通常是需要计算机技术通过算法实现的。然而，算法只能利用已有的信息，而不能凭空创造新的信息，因此使用算法进行解码将要面对从如何从低维数据重建到高维的矛盾^[20,21]。早期主要通过压缩感知理论来解决这一矛盾：将编码-解码过程看作是压缩-解压过程，只不过压缩过程是通过光学系统物理实现的。因此在压缩感知视角下，三维的光谱数据立方体虽然被压缩到二维，但是其信息损失可以通过尽可能优化编码方式而减少，并通过在解码过程中引入变换域的稀疏性，使得通过低维数据刻画和重建高维数据成为可能^[22]。然而在实践中，不仅编码方式由于需要通过光学实现因此难以最优化，并且稀疏变换域也难以通过手工方式找到。因此，压缩感知具有理论上的完备性和优越性，但在实践中却难以达到最优的成像效果。近年来随着人工智能技术的不断发展，神经网络开始提供一个全新的“数据驱动”的视角来重新塑造计算光谱成像的框

架。至少解码的过程有希望能够在“数据驱动”的语境下取代传统的重建过程，并且已经在实验室的小规模测试数据上实现了鲁棒、高质量、高帧率的解码过程^[23]。

以大模型为主要特征的新一代人工智能技术，如 ChatGPT^[24] 和 AlphaFold3^[25]，展现出了惊人的多任务、多模态、复杂场景的处理能力，为传统的神经网络应用提供了指引性的方向。如何将大模型这一强大的工具应用到计算光谱成像领域，以期能提供高性能、高鲁棒性、多模态的计算光谱成像框架，将成为这一领域至关重要的问题。然而训练大模型的最重要的需求之一便是大规模、多样性、高质量的训练数据，这对于计算光谱成像领域而言是难以获取的^[26,27]。

在这一问题背景下，本文尝试在多样化的计算光谱成像编码框架中，选择长波红外多孔径快照式光谱成像^[28,29]、计算微型光谱仪^[30] 和非视距光谱成像^[31] 三种框架，分别根据编码框架的特点，针对性地提出有监督学习、混合监督学习、自监督学习三种思路。其中有监督学习与物理仿真模型结合，在可大规模生成的高质量仿真数据上训练。自监督学习通过物理成像模型和隐式神经表达，不再需要额外训练数据。混合监督学习同时结合了有监督学习和自监督的特点。本文提出的三种学习框架，能够适用于不同的计算光谱成像问题，为新一代人工智能技术在计算光谱成像领域的应用提供思路。

1.2 相关领域研究现状

1.2.1 光谱重建与优化方法

尽管不同的光谱编码调制技术具有不同的物理架构，但由于其压缩测量的固有性质，它们共享一个共同的前向测量模型。这个模型可以表示为以下线性方程组：

$$y = \Phi f + n \quad (1-1)$$

其中， $y \in \mathbb{R}^{MN}$ 表示系统测量值， $f \in \mathbb{R}^{MNL}$ 是大小为 $M \times N \times L$ 的完整数据立方体， $\Phi \in \mathbb{R}^{MN \times MNL}$ 是传感矩阵， $n \in \mathbb{R}^{MN}$ 是系统噪声。

这个求解问题是一个典型的病态问题，通常基于两种方法：通过构建约束条件解决凸优化问题，或使用数据驱动的神经网络方法进行数据预测。

基于优化算法的重建方法直接使用系统的前向模型和手动选择的先验信息来构造优化问题。用于构建优化问题的先验信息通常来自传统上用于图像重建和修复的方法，如稀疏先验^[32]、平滑先验^[33]、低秩先验^[34] 等。稀疏先验假设数据可以在变换域中表示为 $f = \Psi t$ ，其中 f 是数据立方体， Ψ 是稀疏逆变换， t 是数据立方体在稀疏变换域中的表示。根据实际场景的需求，这个变换可以是傅里叶变换、离散余弦变换、小波变换等。由此产生的优化问题可以表示为：

$$\arg \min_t \|y - \Phi \Psi t\|_2^2 + \alpha \|t\|_1 \quad (1-2)$$

其中 $\|\cdot\|_1$ 是 1-范数, 用于约束稀疏性。平滑先验基于观察到自然图像通常具有一定程度的平滑性。优化问题可以表示为:

$$\arg \min_f \|y - \Phi f\|_2^2 + \alpha \|f\|_{TV} \quad (1-3)$$

其中 $\|\cdot\|_{TV}$ 是正则化项总变差 (Total variation, TV) 计算图像中相邻像素的梯度, 是图像重建中常用的平滑度度量。低秩先验基于自然图像往往具有低秩结构的事实, 考虑了内部数据相关性。优化问题可以表示为:

$$\arg \min_f \|y - \Phi f\|_2^2 + \alpha \|f\|_* \quad (1-4)$$

其中 $\|\cdot\|_*$ 是正则化项核范数。

针对这些模型, 研究者提出了多种高效求解算法。Bioucas 等提出的 TwIST (Two-Step Iterative Shrinkage/Thresholding) 算法^[35] 通过改进 IST (Iterative Shrinkage/Thresholding) 的迭代策略, 显著加速非二次正则化问题的收敛过程。Figueiredo 等^[36] 开发的梯度投影方法则解决了带约束的稀疏优化模型, 在计算时间上优于同期方法。Yuan 团队^[37] 将交替方向乘子法 (Alternating Direction Method of Multipliers, ADMM) 与 TV 正则化结合, 构建了针对编码孔径快照系统的 ADMM-TV 框架, 并揭示了其与广义交替投影 (Generalized Alternative Projection, GAP) 的算法关联性。

非局部建模为优化方法提供了新维度。Dong 等^[38] 引入结构稀疏性约束, 提出非局部低秩正则化 (Nonlocal Low-rank Regularization, NLR) 框架, 利用非凸秩代理函数增强细节恢复能力。Liu 等^[39] 通过联合非局部自相似性与秩最小化模型, 显著提升了快照式压缩成像 (Snapshot Compressive Imaging, SCI) 系统的重建精度。字典学习方法为数据自适应优化开辟了新路径: Yuan 等^[40] 设计盲压缩感知算法, 通过 RGB 图像辅助训练耦合字典, 实现了高光谱数据的特征保持重构。Yang 等^[41] 基于最大边际似然估计算法, 直接从压缩测量中学习高斯混合模型, 在实时光谱成像应用中展现出优越性能。

尽管优化方法具备理论完备性, 其实际应用仍面临显著限制。多数算法依赖迭代求解机制, 通常需数分钟计算时间且精度受限。人工设计正则项难以准确刻画场景特性, 导致对复杂目标的泛化能力不足。这些问题催生了神经先验方法的发展, 通过神经网络自动学习图像特征与重建映射, 为光谱重建问题提供新范式。

1.2.2 光谱重建与神经先验方法

基于深度学习方法的新型重建框架针对传统优化方法的不足展开创新, 其核心在于将神经网络学习到的先验知识与传统优化模型相结合。此类方法构建的优化问题可表示为:

$$f = \arg \min_f \|y - \Phi f\|_2^2 + \alpha R(f) \quad (1-5)$$

其中 $R(f)$ 为深度先验正则项，通常通过在外部光谱数据集预训练或在优化过程中同步训练获得。此范式在保持模型驱动优势的同时，显著增强了先验建模的适应性。

Huang 等人^[42] 提出结合深度高斯尺度混合 (Gaussian Scale Mixture, GSM) 先验的最大后验估计框架，通过端到端联合优化实现光谱压缩成像的精细重建。该方法集成 Swin Transformer 捕捉长程依赖关系，在仿真与真实实验中均超越传统先验方法与监督学习模型，验证了可解释性与性能的协同提升机制。Fu 等^[43] 设计基于双层学习的卷积神经网络架构，创新性地集成外部预训练与单样本自适应机制，通过光谱通道注意力与空间-光谱联合约束提升重建保真度，在编码高光谱成像任务中实现跨系统性能优化。

Chen 等^[44] 提出融合低秩子空间分解与数据驱动的光谱重建方法，通过分离全局谱相关特征并构建空间注意力网络自学习系数先验。该方法突破传统模型对人工特征的依赖，在无监督条件下通过交替优化实现重建精度提升，多模态成像验证显示其信噪比平均提升 2.8 dB。Ma 等^[45] 研发深度张量 ADMM-Net 框架，将张量优化迭代展开为可学习网络层结构，自适应提取低秩特征实现快速视频重建，较传统方法实现 0.8 至 2.5 dB 的信噪比提升及 1500 倍加速效果。

Meng 等^[46] 提出基于未训练网络的自监督光谱重建框架，将深度图像先验 (Deep Image Prior, DIP) 嵌入即插即用优化过程，无需外部数据即可达到监督学习相当的重建性能。该框架通过集成预训练去噪先验进一步改善质量，为无校准实时光谱成像建立新范式。

可插即用 (Plug-and-play, PnP) 优化框架构成另一重要分支，其核心在于用深度去噪器替代传统优化的迭代去噪步骤。Xie 等^[47] 提出多镜头光谱成像 PnP 方法，结合加权低秩先验与非局部自相似性构建新型去噪器，在双镜头压缩采样下实现超 50 dB 峰值信噪比。Zheng 等^[48] 设计深度去噪先验模块嵌入 PnP 框架，在五类光谱成像系统上验证单次快照多光谱重建能力，为实时动态成像提供技术路径。Chen 等^[49] 提出 LR2DP 方法，通过低秩基空间投影与双深度先验协同机制，在半二次分裂算法框架下实现信噪比提升 36.8%。

神经先验框架的优势体现在两方面：(1) 预训练网络特性可将运算效率提升两个数量级；(2) 对编码掩模与采样率变化的鲁棒性保障了系统灵活性。研究表明，当结合 ADMM/GAP 算法与高效去噪器时，其重构质量可达前沿方法水平，在速度-精度-灵活性间达成理想平衡。但该框架在推广应用时仍面临挑战：一方面，高光谱等特殊成像系统缺乏适配的灵活去噪器，需针对性开发新型网络架构；另一方面，最终性能对去噪器质量的高敏感度导致模型泛化能力受限，这构成当前研究的主要技术瓶颈。

1.2.3 光谱重建与神经网络方法

基于优化的光谱图像重建方法通常需要在迭代过程中花费长时间的重建时间。此外，手工设计的先验知识对自然光谱数据的描述并不精确，导致空间和光谱重建的精度并不稳定。例如，全变分先验总是用于重建算法，但有时会对结果产生过度平滑。

神经网络技术可以应用于光谱成像过程中的每个步骤，如光学调制策略的设计或者是找到具有代表性的正则化项，整个重建过程都可以用神经网络替代。采用神经网络方法可以将重建速度提高数百倍。此外，通过神经网络从大量光谱数据中学习先验，可以提高空间和光谱重建的精度。基于神经网络使用的不同位置，本文将基于神经网络的压缩重建方法分为四类：(1) 有监督学习，使用深度神经网络进行直接重建；(2) 无监督学习，使用神经网络的广泛范围作为先验，并进行迭代重建。(3) 联合设计学习，同时学习光学调制设计模式和后续的重建网络；

1.2.3.1 有监督学习

有监督学习框架下的光谱重建任务通常采用端到端神经网络架构，将压缩测量直接映射为高维光谱数据重构结果。此类方法的核心在于通过大量训练样本学习从压缩域到原始光谱空间的非线性映射函数，其性能高度依赖于网络结构设计、损失函数选择以及训练策略优化。

网络架构创新方面，研究者广泛借鉴计算机视觉领域的先进模型，包括引入残差连接增强梯度传播^[50]、构建密集连接促进特征复用^[51]、以及嵌入注意力机制提升谱-空域建模能力^[52]。Jiang 等^[53]开发的单幅高光谱超分辨率网络结合空间-光谱先验网络与渐进式上采样框架，设计具有光谱注意力机制的空间-光谱残差模块，在多个数据集上将峰值信噪比 (Peak Signal Noise Ratio, PSNR) 指标提升 2.1 dB 以上。Miao 团队提出的 λ -Net^[54] 基于条件生成对抗网络架构，通过自注意力机制与分层重建策略，将高光谱重构时间从传统方法的数小时缩短至亚秒级，峰值信噪比提升达 4 dB。

损失函数设计逐步突破像素级差异的局限，融合物理驱动与语义感知约束。主流方法除采用反投影像素损失确保数据保真度^[55]外，还引入基于特征匹配的感知损失^[56]与风格损失^[57]，通过预训练 VGG (Visual Geometry Group) 网络提取多尺度特征，增强重建结果的感知一致性与纹理细节。

深度展开网络作为结合物理模型与数据驱动优势的代表性范式，展现出显著性能优势。Wang 等^[58]提出的 GAP-CCoT (Convolution and Contextual Transformer) 网络将卷积-Transformer 混合模块嵌入广义交替投影框架，实现重建质量 PSNR 提升 2 dB 且运算效率优化 80%。Qin 团队构建的增强型深度展开网络 (Enhanced Deep unrolling Net, EDUNet)^[59]通过跨阶段谱自注意力机制与光谱几何一致性损失函数，在 KAIST (Korea Advanced Institute of Science and Technology) 等基准数据集上全面超越 15 种对比算法，四项关键指标平均提升 18.7%。Meng 等^[60]

设计的 GAP-net 采用逐阶段可训练 CNN 去噪模块，凭借信号调制泛化能力在真实数据集达到 3.5 dB 重构增益，并通过可证收敛性为物理嵌入网络提供理论支撑。

针对动态光谱成像的特殊需求，Cheng 等^[61]提出双向循环网络，通过时序帧传播与谱-空域联合优化路径，实现视频/高光谱重建 PSNR 提升 3 dB。其扩展架构彩色双向循环网络更通过解压缩-去马赛克一体化设计，在真实多光谱系统达成 90.2% 色彩还原精度。

尽管监督学习方法已实现亚秒级的高质量重建，其性能仍受限于对大规模标注数据的依赖。当前研究正通过数据增强、物理约束正则化与小样本学习等策略缓解该问题，但泛化能力与复杂场景适应性仍是核心挑战。

1.2.3.2 无监督学习

无监督学习框架为计算光谱重建提供了无需标注数据的替代解决方案，其核心思路在于通过设计网络架构与损失函数，将光学成像的物理前向模型与数据分布先验隐式嵌入优化过程。经典方法如深度图像先验^[62]通过未训练卷积网络的结构约束（如感受野特性与频谱衰减），既可覆盖常见光谱图像的解空间，又避免了模型对训练集的过拟合风险。尽管此类方法依赖迭代优化过程牺牲了实时性，但其对新型光谱成像系统的自适应能力显示出独特优势。

Sun 等人提出的 HCS2-Net^[63]通过可微分渲染层融合物理模型，构建编码孔径参数自适应的端到端生成对抗网络。其创新性体现在：(1) 引入空间-光谱双层注意力机制联合建模局部谱线关联性；(2) 设计融合复数感知损失的对抗训练策略，增强重建结果的光谱保真度。实验表明，该方法仅依赖单次快照即可灵活适配多光谱系统特性，在 CAVE、Harvard 等数据集上实现平均 2.3 dB 的 PSNR 增益，为无监督光谱重建建立了新基准。

Bacca 团队提出的 Tucker 低秩嵌入网络^[64]通过张量分解建模光谱结构特性，其核心创新是将 Tucker 低秩约束等价转换为神经网络层拓扑结构：(1) 输入层执行光谱维度投影以获取核心张量；(2) 耦合网络模块迭代优化模态因子矩阵；(3) 重构层基于张量积恢复数据立方体。该方法通过最小化测量残差的 L_2 约束，在真实编码孔径系统中取得与数据驱动方法相仿的重建精度，验证了模型驱动的有效性。

Kim 等人在非负张量分解领域的奠基性研究^[65]为光谱重建提供了理论支撑。其提出的网络化张量低秩层通过约束核心张量的稀疏度，实现光谱数据全局-局部结构的自适应分解，推动了后续基于张量代数的深度学习架构发展。

Liu 等人提出的监督-自监督混合框架^[66]通过两阶段学习机制突破无监督建模瓶颈：第一阶段在外部数据集预训练网络提取光谱形态先验；第二阶段采用 Stein 无偏风险估计与噪声方差自适应机制，在目标域实现自监督微调。该方法在多个成像模态中的实验表明，其 PSNR 指标比纯无监督方法提升 1.8-2.5 dB，且对 25-50 dB 噪声环境表现出显著鲁棒性。

总体而言，无监督学习在光谱重建中的核心优势体现为对硬件参数波动性与场景多样性的强鲁棒性，但其迭代驱动的重建效率（通常需数分钟）仍是部署瓶颈。当前研究通过引入物理引导的神经网络架构、可微渲染层加速算子等手段提升计算效率，在保持零样本适应性的同时将重建时间缩短至秒级，展现出较高的技术实用价值。但如何在光学系统噪声建模、动态场景适应性等层面实现突破，仍是该方向亟待解决的关键挑战。

1.2.3.3 联合设计学习

传统方法中，前向测量模型设计多基于压缩感知理论，借助随机模式维持受限等距特性（Restricted Isometry Property, RIP）^[67]。近年来，联合设计学习框架突破光学端与算法端的孤岛优化范式，其核心在于将物理调制机制参数化为可微分网络组件，通过端到端梯度反传实现系统级优化。

现有研究主要通过两类方式实现光学组件的可学习设计：(1) 直接将二元/多级调制模式参数化为网络权重；(2) 构建编码结构生成网络，通过条件生成过程输出物理实现的调制模式。Arguello 团队创新性地将秩最小化约束引入多帧 CASSI 系统优化^[68]，通过光谱选择性编码组合增强压缩感知特性。其后续工作提出色化编码孔径优化框架，在离散化光强分布设计中突破传统二值编码的光谱复用限制，实现测量次数减少与 PSNR 增益的协同优化。

Wang 等提出的 HyperReconNet^[69] 开创了编码孔径与重建网络端到端联合训练的先河。该方法通过空谱联合建模网络将重复编码模式融入物理层设计，在仿真实验中实现重建精度与硬件适应性的显著提升。Galvis 研究组通过超分辨率算法与合成编码的协同设计^[70]，有效消除微反射镜阵列（Digital Micromirror Device, DMD）像素失配引起的空间混叠效应，在真实 CASSI 系统中提升 22% 空间分辨率。其扩展研究创新性地引入 RGB 辅助信息指导高频编码生成，通过边缘特征增强策略实现 3 dB 的信噪比增益。

多维光调制成为新一代研究热点，Lin 等提出的双共轭位置调制框架^[71]通过光机协同设计实现单帧高光谱捕获。Shi 等人构建的多通道透镜阵列-滤光片联合优化系统^[72]，在 430-700nm 谱段实现 31 通道高质量重建，PSNR 指标较传统方案提升 5 dB 以上。Hinojosa 等^[73]提出面向聚类任务的编码优化方法，通过 3D-SSC 算法直接在压缩域保留判别特征，实现与全数据聚类误差 <4% 的突破性成果。

在硬件实现方面，硬件参数的梯度回传依赖近似梯度估计策略。BinaryConnect 方法^[74]通过量化感知训练将浮点权重映射至物理调制模式，有效解决了离散调制不可微问题。Zhang 等提出的深度学习驱动宽带编码系统^[75]采用自适应滤光片与神经网络协同设计，在算法端实现 5nm 光谱分辨率重建的同时，硬件端处理速度提升 7000 倍。当前技术瓶颈集中体现在光电调制器件的动态响应建模精度、复杂光路参数的联合优化效率，以及多物理场耦合的可微仿真能力等方面。

联合学习设计通过将光学编码孔径参数化嵌入深度学习网络，实现了光谱压缩感知与重建算法的端到端协同优化，显著提升了系统的自适应性和重构精度。该框架利用可微分建模与梯度反传机制，在突破传统随机编码的物理层局限性的同时，通过任务驱动的联合优化范式为高动态、多场景光谱成像系统提供了高效解决方案。

1.2.3.4 像素级光谱重建

早期的波长编码光谱重建工作是在 RGB 图像上进行像素级的重建。他们考虑了如何从一个 3 通道的 RGB 向量重建一个具有更多通道的光谱向量的减少问题。像素级光谱重建使得神经网络不需要大量的光谱图像数据集进行训练，而是直接在像素光谱上进行训练，因此数据集的规模得以大大提升。一般来说，这些像素级方法寻求单个光谱的表示，并开发了从该表示重建光谱的方法。这些像素级光谱重建的任务通常通过寻求隐藏的流形嵌入空间以有效表达光谱的光谱流形学习来实现。

流形学习假设一个低维流形嵌入在高维数据空间中，并试图从高维空间中的数据恢复低维流形。Jia 等人^[76]提出了一种三步方法进行流行学习重建：(1) 通过等距特征映射 (Isomap^[77]) 找到流形空间的适当维度；(2) 训练一个径向基函数 (Radial Basis Function, RBF) 网络将 RGB 向量嵌入到低维流形中，这决定了成像模型前向映射 Φ 的逆；(3) 使用字典学习将流形表示映射回光谱空间，这决定了成像模型的前向映射 Φ 。其中的 RBF 网络和字典学习方法可以被深度神经网络（如 AutoEncoder）替代以提高性能，因此可以进一步提升基于流形的重建。

1.3 光谱重建存在的问题与挑战

研究现状揭示，神经网络的光谱重建方法作为速度快、能获取高质量光谱数据的手段，已经实现了显著的进步，并已成为这个领域的重要部分。然而，依然面临几大挑战，如训练数据的不足、光谱分辨率受编码设计的限制以及全局前向测量模型的重建问题等待解决。

(1) 训练数据缺乏。神经网络方法中，训练数据的高质量是对重建效果至关重要的。但与计算机视觉中使用的通用设备，如通用 RGB 相机不同，光谱测量器具有众多种类，产生的光谱图像存在明显差异，尤其是波长范围和光谱分辨率的不同，使得各设备之间的数据难以通用。因此，通常需要定制数据采集，方可满足每个光谱重建的训练数据需求，这对仪器研发和应用团队通常是一项耗时且复杂的任务。

(2) 编码设计受制于光谱分辨率。因物理光学硬件系统的局限，光谱成像的编码设计需服从物理约束，此也限制了前向测量模型，无法达到理论上的最优编码设计。在某些特定的光谱成像系统中，例如多孔径分光光谱成像系统或微型计算光谱仪，由于其自身对实时性或设备体积大小的需求存在限制，导致编码设计

更须妥协，前向测量模型的效度比传统光谱成像系统更低，光谱重建任务更具挑战，系统的光谱分辨率变得较低。

(3) 全局性前向测量模型的重建仍待解决。在实际光学系统中，由于硬件系统的限制，光谱成像的前向模型通常呈局部性，例如，探测器上某个像素接收到的光学信号仅与其邻近像素有关。这个特性通常可用于减小光谱重建的运算量和空间占用，提高算法的运行速度。然而，对一些特殊场景光谱成像任务，例如非视距条件下的光谱成像，这个局部假设不再适用。因此，亟需新的方法来应对全局光谱成像问题。

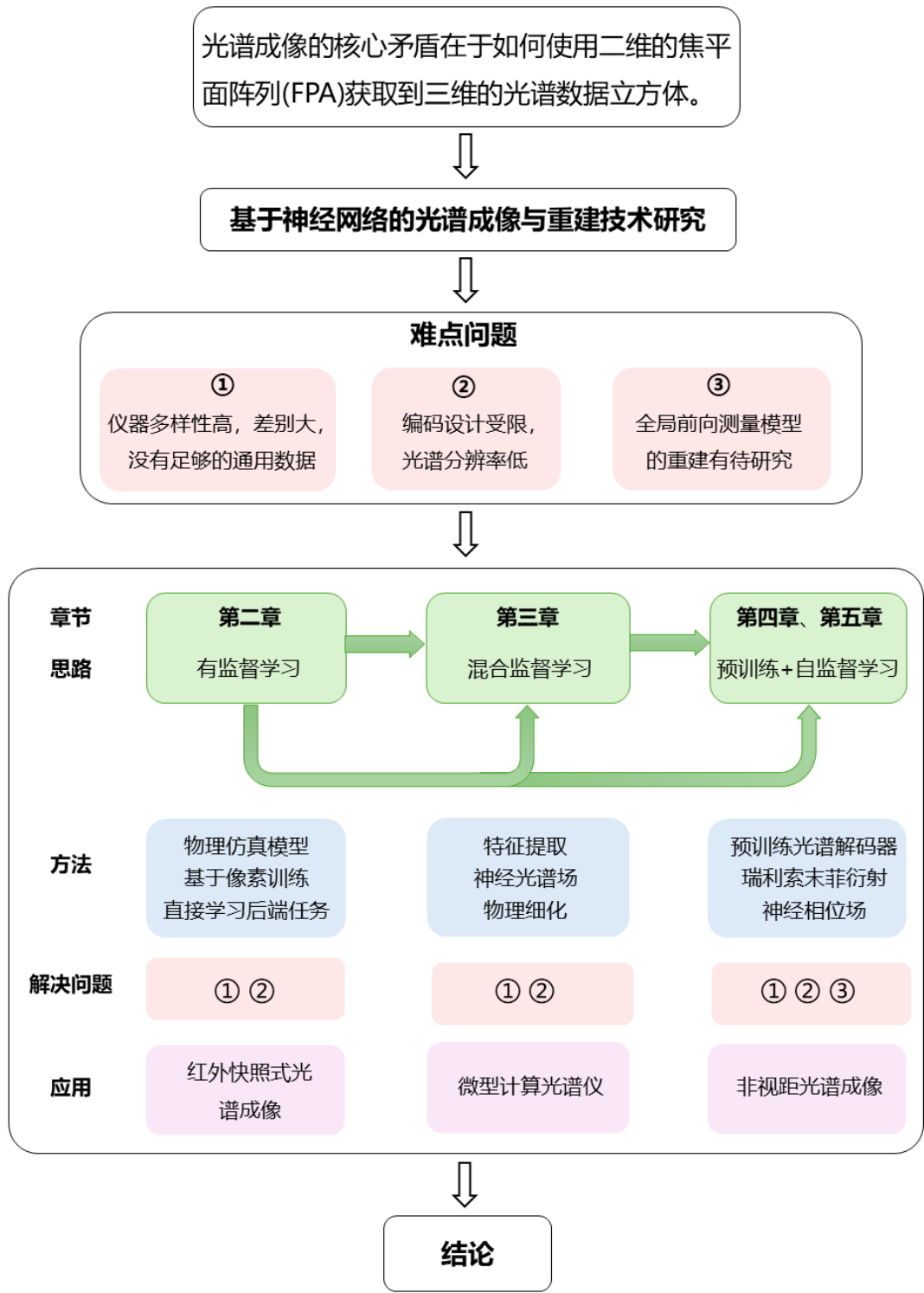


图 1-1 论文研究内容框架
Figure 1-1 Pipeline

1.4 本文主要研究内容

本研究主要尝试解决人工智能与神经网络在处理计算光谱成像问题时所遭遇的挑战。尽管神经网络近年来在解决成像问题方面展现出了强大的能力，但仍然面对通用数据稀缺、光谱分辨率受编码设计限制以及缺乏全局前向测量模型下的重建方法等问题。因此，根据是否需要额外训练数据的分类方法，本文提出了三种不同的学习框架：基于物理仿真模型的有监督学习，基于神经光谱场的混合监督学习，以及基于预训练和光谱相位场的自监督学习。这三种框架可适应不同的计算光谱成像问题，为人工智能在该领域的进一步应用提供了可能的解决方案。

针对通用数据稀缺和光谱分辨率受编码设计限制的问题，本文首先提出了基于物理仿真模型的有监督学习方法。在众多光谱成像场景中，尤其是红外光谱成像任务中，这两大问题常常严重削弱了成像质量。为应对这些挑战，本文提出了一种新颖的神经网络方法，基于像素的层次化光谱变换（Hierarchical Spectral transformer, HST）框架。此框架是一种从公开的单像素红外光谱数据库中学习而来的深度学习模型，旨在实现红外光谱图像的高光谱分辨率重建，从而加强其在局部和全局上下文中对光谱数据的理解。此外，本文还提出了 Cross-talk transformer，以直接在低质量数据上执行高鲁棒性的后端任务算法，以提升长波红外在气体识别领域的应用效果。实验证明，这两种方法均在提升光谱分辨率、降低噪声以及监测气体类型和浓度方面表现出高精度和强大稳定性。

在此基础上，本文进一步提出了一种基于神经光谱场的混合监督学习策略，用以进行光谱重建。该策略首先利用物理模型模拟的数据进行有监督学习，然后通过神经光谱场的建模方式优化神经网络输出的光谱曲线，最后引入一种基于物理的细化方法，以增强重构光谱与测量数据的一致性，进而实现高质量的光谱重建。测试结果表明，这种策略可以有效提升微型计算光谱仪的波长精度和光谱分辨率，且在成像性能测试中也展现出优秀的光谱成像能力。

针对全局前向测量的光谱重建问题，本文最后提出了一种基于预训练和光谱相位场的自监督学习框架。该框架通过引导神经网络的预训练来优化光谱解码器，以实现光谱信息的精确解析，同时引入光谱相位场的概念，并将其纳入神经网络的训练中，以降低噪声的干扰，提高光谱成像的质量。实验结果表明，该框架在非视距光谱成像环境中能有效缩短采集时间，提高重建质量，显著提升物质鉴别能力，为隐蔽场景的非侵入式分析提供了实用化解决方案。

通过以上研究，本文系统性地探讨了神经网络在光谱成像领域中的应用，深入研究了计算光谱成像的编码-解码框架，并在此基础上分别提出了基于有监督学习、混合监督学习和自监督学习的学习策略。这些策略在长波红外多孔径快照式光谱成像、计算微型光谱仪和非视距光谱成像等典型场景中得到了有效应用，取得了优于以往的重建结果。

1.5 本文章节安排

本文第一章为绪论部分，主要阐述了光谱成像问题的形成和重要性以及用神经网络解决问题的优势和局限。根据这一基础，本文分别介绍了两种解决光谱成像问题的主要方法，即基于优化和基于神经网络的方法，并进一步分析了当前光谱重建领域存在的问题和挑战。

第二章重点介绍了基于物理仿真模型的有监督学习光谱重建方法，包含两个具体的实施方案：一是应用人工智能重建高光谱图像，并提出了一种名为 **Hierarchical Spectral Transformer (HST)** 的基于像素的方法，能有效提升光谱分辨率并降低噪声；二是运用具有高鲁棒性的后端任务算法处理低质量数据，提出了 **cross-talk transformer** 和辐射传输模型，以产生大量的成像数据并进行高精度的识别和预测。这一方法特别适应各类红外光谱成像系统，有助于处理仪器多样性高和编码设计复杂等问题。本文在长波红外多孔径快照式光谱成像系统中应用了本文的光谱重建方法，然后这两个方案都证明了其准确的光谱重建和气体监测能力，以及其在各种应用中的潜力和有效性。

第三章突出介绍了基于神经光谱场的混合监督学习光谱重建方法。首先，本文利用物理模型模拟得出的数据进行有监督学习，然后通过神经光谱场的建模进一步优化神经网络输出的光谱曲线。最后，本文提出一种基于物理的细化方法，以提高重建光谱与测量数据的一致性，从而实现高质量光谱重建。本文在设计高度受限的微型计算光谱仪上验证了本文的方法，分析了渐变带隙 PN 结光谱仪的光谱响应特性，并在实测的单峰、双峰、精密光谱和自然光谱数据测试验证了所提出方法的有效性，经测试，提出的方法实现了在微米尺寸的光谱仪上达到 10nm 的光谱分辨率，并且实现了远超传统优化重建方法的鲁棒性。

第四章主要介绍了基于神经瞬态场和自监督学习的多光谱重建方法，针对全局前向测量模型的重建问题。由神经瞬态场拓展到多光谱神经瞬态场，本文提出了一个基于隐式神经表达的多光谱神经瞬态场 (**Multi Spectral Neural Transient Fields, MSNeTF**) 成像框架，适应全局前向测量模型的重建流程。本文在非视距成像条件下验证了本文提出的框架，实现了非视距多光谱成像的神经网络方法，在合成和真实数据集上的实验表明，**MSNeTF** 实现了最先进的性能。

第五章主要讨论了基于预训练和光谱相位场的自监督学习光谱重建方法。此框架综合考虑了光谱重建问题中数据量、光谱分辨率以及全局测量重建等挑战，采用了先在大规模光谱数据库中预训练光谱解码器，再将解码器应用于自监督学习的光谱相位场的策略。本文在非视距条件下测试了提出的框架，分析了非视距成像的物理模型和成像机理，进一步分析了非视距光谱成像中的训练数据有限和采集数据限制问题。在仿真和实测数据上验证提出的方法，证明能够在非视距条件下实现 10nm 的光谱分辨率成像和近 30 倍的数据需求降低。

第六章总结了本研究的成果和存在的不足，并对基于神经网络的光谱成像系统的未来发展与研究方向作出分析。

第2章 基于物理仿真模型的有监督学习光谱重建

2.1 引言

计算光谱成像在许多领域已发挥了关键作用。然而,在许多场合,尤其是红外光谱成像任务中,存在光谱分辨率低和数据噪声大的问题,这大大削弱了成像质量。近年来,神经网络成为解决这些问题的有力工具,因为它可以引入光谱先验知识以提高光谱分辨率,并且在高光谱成像任务中取得了突出成就。然而,由于红外高光谱图像数据严重不足,神经网络在红外光谱成像任务中的应用受到制约,这限制了建立鲁棒模型的可能。同时,现有依赖于卷积层或注意力机制的网络难以有效捕捉到局部以及全局的光谱相关信息。

为了解决这些限制,一个思路是利用神经网络实现高光谱图像的重建。本文推出一种新的神经网络方法,基于像素的 Hierarchical Spectral Transformer 框架,是一种从公开的单像素红外光谱数据库中学习而来的新颖深度学习模型。HST 旨在实现红外光谱图像的高光谱分辨率重建,从而加强其在局部和全局上下文中对光谱数据的理解。经过在仿真和实测的红外光谱图像数据上的评估,证实了 HST 即使在数据有限的场景中,也能有效提高光谱分辨率和降低噪声方面的鲁棒性。

另一个思路是直接在低质量数据上执行高鲁棒性的后端任务算法。对此,本文提出了 cross-talk transformer,针对长波红外在气体识别领域的应用。该算法利用注意力机制学习目标气体和气体光谱库间的深层相关性。而为了在非制冷快照式红外光谱成像仪(Uncooled Snapshot Infrared Spectrometer,USIRS)上生成大量气体成像数据,本文开发了一个全新的气体辐射传输模型,以充分训练神经网络。实验证明,本文的方法在11种化学气体的实验数据中达到了98.63%的识别精度。在浓度预测上,本文的方法对浓度高达30,000 ppm的化学气体的平均误差仅330 ppm。这些结果突出显示了本文的方法在监测气体类型和浓度方面的高精度和强大稳定性,证实了其在工业监测等场合的潜在应用价值。

2.2 快照式光谱成像系统

2.2.1 多孔径快照式光谱成像

快照式光谱成像的魅力在于能够在单次“快照”中捕获场景的完整光谱信息,从而实现实时动态观察。理想的快照式光谱成像系统应在保持紧凑形态和计算效率的同时,提供高空间和光谱分辨率。然而,实现这一理想需要在空间分辨率、光谱分辨率、成像帧频和硬件复杂度之间进行权衡。其中多孔径分割光谱成像能够保持与商业相机有竞争力的成像帧频,因此非常适用于动态光谱成像。

多孔径分割光谱成像通常使用透镜阵列实现空间复制,结合多光谱滤光片在单次快照内实现光谱立方体测量。Hirai 等人使用这种方法设计了一种多图像

傅里叶变换光谱仪^[78]。该光谱仪利用带倾斜镜子的迈克尔逊干涉仪在探测器上生成与位置相关的光程差，用以产生干涉子图像。随后通过建立干涉图和光谱之间的傅里叶变换关系来实现三维光谱重建。类似地，Kudenov 等人用双折射 Nomarski 棱镜替代迈克尔逊干涉仪，构建了一个更紧凑且抗振动的快照光谱仪^[79]。棱镜相对于探测器的旋转角度经过精心设计，使子图像具有不同的光程差。随后通过傅里叶变换实现光谱重建。Hubold 等人开创了一种结合倾斜线性渐变滤光片和透镜阵列的成像方案^[80]。该方法不仅降低了制造难度和成本，还引入了由于线性渐变滤光片倾斜角度导致的探测器浪费和光谱扰动之间平衡的挑战。2019 年，Mu 等人引入了一种通过倾斜透镜阵列而不是线性滤光片的改进方法，显著减少了光谱扰动，并确保探测器利用的稳定性^[81]。这两种方法实现连续光谱的同时多次采样的机制非常相似。复制的各个空间子图像由不同的连续光谱扫描，因滤光片和探测器之间的角度而产生完整的空间光谱数据。然而，透镜阵列的尺寸决定了系统能够实现的最大的光谱通道数，空间维度的复制意味着空间维度像素数的平均分布，光谱分辨率的增加必然导致空间分辨率的下降。2024 年，Yang 等人首次提出将多孔径分割光谱成像技术推广到长波红外波段，并针对气体探测进行了光学设计和电子学系统优化^[82]。多孔径快照式光谱成像能够达到高帧频的实时光谱成像，但光谱分辨率低，因此如何在解码过程中实现高光谱重建和后续的应用仍然是一个挑战。

2.2.2 长波红外多孔径快照式光谱成像系统

长波红外（Long-wave Infrared, LWIR）光谱成像在气体监测、矿产勘探和火灾检测等各种应用中发挥着关键作用。近年来，借助先进算法的计算光谱成像技术取得了显著进展，使得高质量光谱图像能够实时获取，比如使用 USIRS 系统。然而，USIRS 系统面临一些挑战，尤其是光谱分辨率低和数据噪声大，这些问题可能会降低图像质量。深度学习已成为解决这些挑战的一种有前途的方案，因为它在处理噪声数据方面特别有效，并且在高光谱成像任务中取得了显著成功。然而，深度学习在长波红外光谱成像中的应用受到了长波高光谱图像数据严重缺乏的制约，这限制了鲁棒模型的训练。此外，依赖卷积层或注意机制的现有网络难以有效捕捉局部和全局的光谱相关性。

多孔径快照式光谱成像系统的原理如图 2-1(a) 所示。在多孔径系统中，核心成像组件是滤光片和透镜阵列。滤光阵列用于划分光谱范围。每个子滤光器是一个窄带滤光器，它决定了子孔径的光谱范围。透镜阵列中的子透镜对相同的场景进行成像，起到复制的作用。对每一个子孔径，透镜阵列的子透镜与滤光阵列的子滤光器一一对应。最终实现了在单次曝光时间内获得了不同波段下同一场景的多张图像。

气体光谱成像系统捕捉的数据包括一维光谱信息（图中画出六个通道）和气体场景的二维空间信息。如图 2-1(b) 所示，焦平面块的不同颜色代表不同的波段。根据红外波段的气体吸收特性，每种气体在每个波段有独特的吸收强度。因此，在不同光谱带的图像中，气体区域具有相同的空间信息和不同的光谱信息，

这些信息可以通过算法进一步转换为气体的类型、形状和浓度信息。

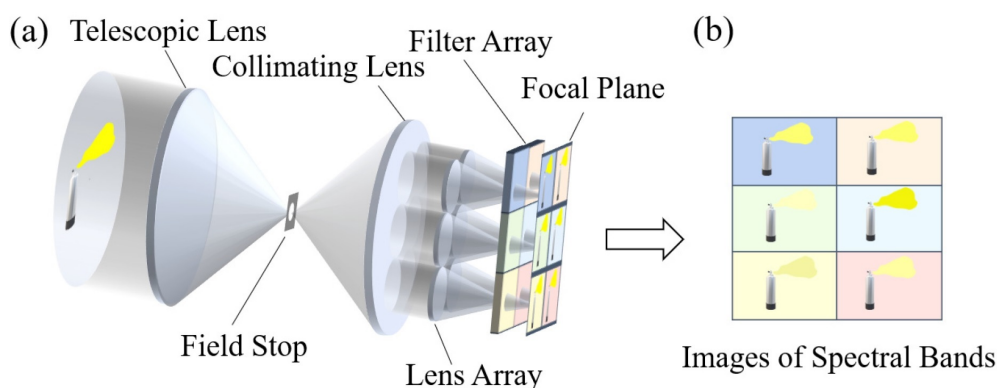


图 2-1 多孔径快照式光谱成像系统原理

Figure 2-1 Principle of Multi-Aperture Snapshot Spectral Imaging System

非制冷快照红外光谱成像是根据多孔径快照式光谱成像系统原理开发的工业仪器，如图 2-2(a) 所示。USIRS 使用滤光器和多孔径成像技术在各个波长下捕捉图像，从而提供目标物体的详细光谱图像。图 2-2(b) 展示了 USIRS 的光学设计。整个过程从一个望远镜镜头收集目标物体的光线开始，这些光线随后被导向一个准直镜头，确保光线平行。接着，由九个小镜头组成的透镜阵列再现了目标场景。光线然后通过一个滤光器阵列，其中每个滤光器对应一个特定的波段，最终到达焦平面，在那里生成并捕捉不同波长的图像。

USIRS 利用两个相同的光学系统，唯一的区别是滤光器阵列的波长。这实际上使它成为一个具有 17 个带通波长的双系统：7.22 μm ，7.67 μm ，7.96 μm ，8.33 μm ，8.70 μm ，9.07 μm ，9.48 μm ，9.81 μm ，10.18 μm ，10.55 μm ，10.92 μm ，11.29 μm ，11.66 μm ，12.03 μm ，12.40 μm ，12.77 μm 和 13.14 μm 。图 2-1(c) 展示了 USIRS 的一个图像，目标场景包括两个个体和两个分别释放氨气 (NH_3) 和六氟化硫 (SF_6) 的气瓶。图 2-1(d) 展示了蓝色框标出的气体的光谱曲线。由于 USIRS 只有 17 个带通光谱通道，且测量通常比较嘈杂，因此它的光谱分辨率不高。为了进一步应用，有必要对其进行去噪并重建高分辨率光谱。

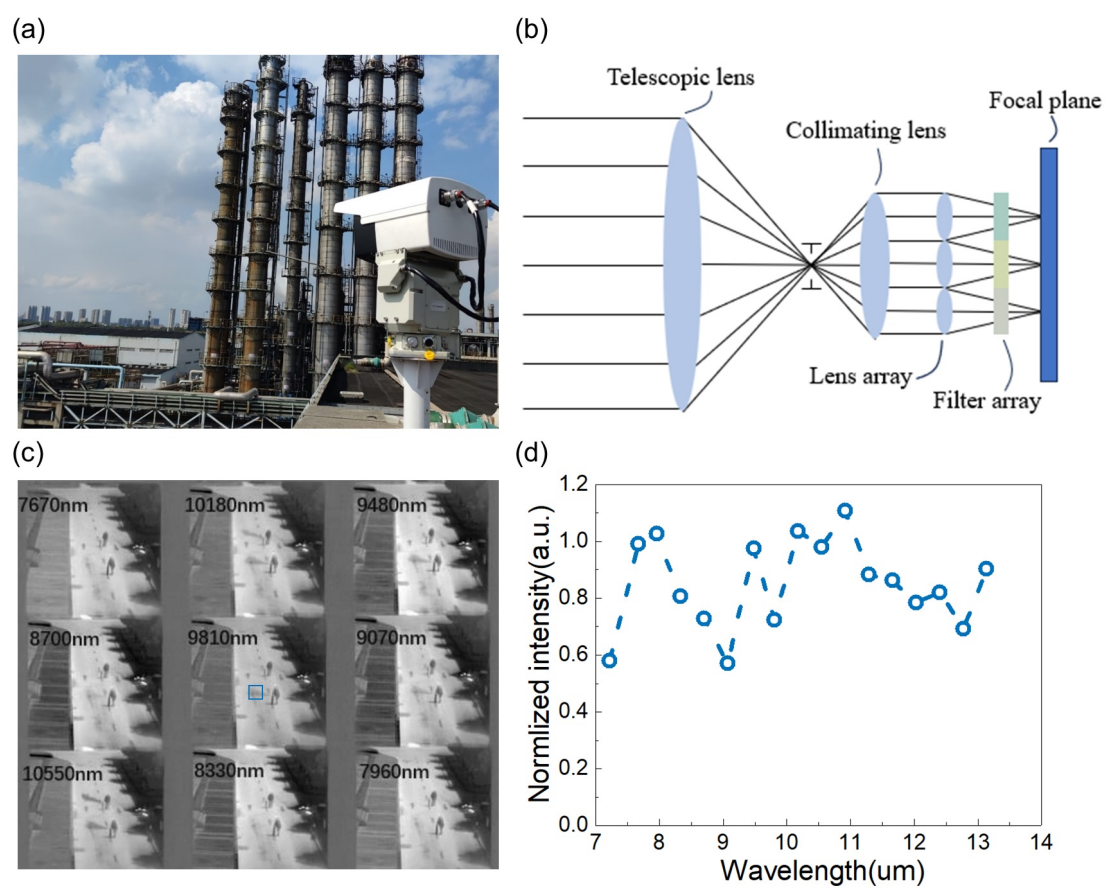


图 2-2 非制冷快照红外光谱仪设计与成像效果

Figure 2-2 Design and Imaging Performance of Uncooled Snapshot Infrared Spectrometer

2.3 像素级长波红外光谱图像重建

解决长波红外高光谱成像挑战的第一个思路是利用人工智能实现高光谱图像的重建。为此本文提出了基于像素的 Hierarchical Spectral Transformer，能够从公开的单像素长波红外光谱数据库中学习的新型深度学习架构。HST 旨在实现高光谱分辨率的长波红外光谱图像重建，增强对光谱数据的局部和全局上下文理解。

2.3.1 信号传输与噪声模型

深度学习是一个有希望解决高噪声高光谱重建问题的解决方案。然而，使用神经网络进行长波红外光谱视频重建的挑战在于缺乏足够的高质量训练数据。现有的大多数长波红外高光谱数据集由光谱仪收集的单点光谱数据组成。因此，本文提出开发 USIRS 的成像和噪声模型，使用此模型从高分辨率数据生成相应的低分辨率数据，并随后在像素光谱级别的数据上训练神经网络。

USIRS 接收到的信号是目标辐射和背景辐射的组合：

$$L(\lambda) = L_t(\lambda)\tau(\lambda) + L_p(\lambda) \quad (2-1)$$

这里， $L_t(\lambda)$ 代表了目标在 USIRS 透镜上的辐射光谱。 $L(\lambda)$ 是到达探测器的辐射， $\tau(\lambda)$ 是成像系统的光谱响应， $L_p(\lambda)$ 是背景辐射。到达探测器的辐射被转换成电流，然后被量化为数字量 (Digital number, DN)：

$$DN = \mathcal{P}(\phi L) + N_d + N_q \quad (2-2)$$

在这个方程中， $\mathcal{P}$ 是光子噪声算子， ϕ 是探测器的量子效率， N_d 是电路中的暗电流， N_q 是量化噪声。暗电流 N_d 和量化噪声 N_q 都假设服从高斯分布，分别为 $\mathcal{N}(\mu_d, \sigma_d)$ 和 $\mathcal{N}(\mu_q, \sigma_q)$ 。

成像和噪声模型提供了将公开的高分辨率光谱数据转换为 USIRS 合成测量数据的管道，从而为神经网络训练提供配对数据。

2.3.2 网络结构

本文提出一种全新的模型 Hierarchical Spectral Transformer，其采用自注意力机制捕捉长距离依赖和频谱相关性，满足现实场景中对全局信息的要求。

HST 采用多阶段架构逐步细化光谱和空间特征，从而显著提高重建精度，这一点对于实际应用至关重要。模型设计充分保留全局与局部光谱信息，有助于提升长波红外光谱图像重建性能，实验结果证明了这一优势。

图2-3中详细展示了 HST 网络的各组成模块与整体结构。

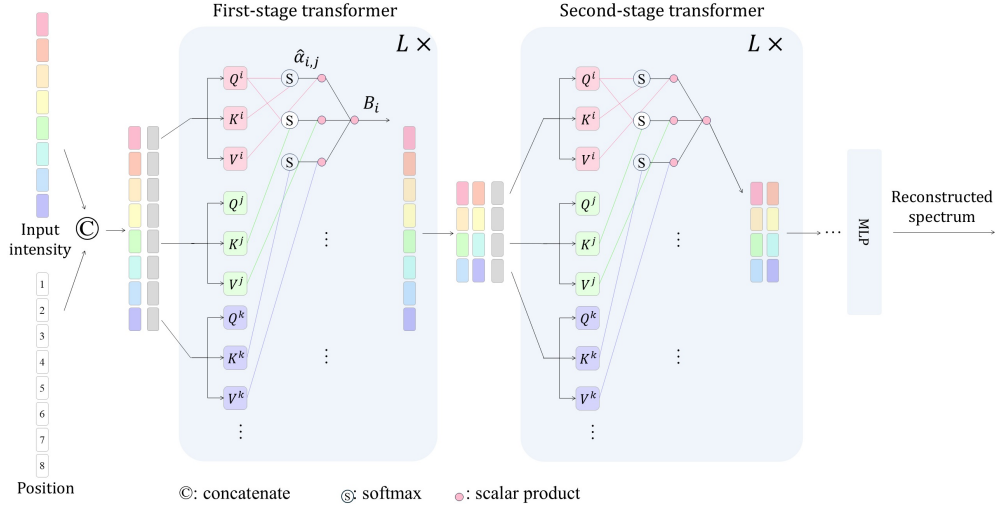


图 2-3 像素级分层光谱 Transformer 结构示意图

Figure 2-3 Schematic Diagram of Pixel-Level Layered Spectral Transformer Structure

2.3.3 位置编码

在 HST 网络设计中，输入为现实测量获得的强度信息 $I \in \mathbb{R}^I$ 。由于不同光谱通道具有各自的物理属性，强度 I 在各通道中表现出差异性。

为充分融合位置信息，引入位置编码 $Pos \in \mathbb{R}^{2M^I}$ 。对每个特征位置 m ，采用三角函数转换生成位置嵌入，其表达式为：

$$Pos(m) = [\sin(m \times 2^0), \cos(m \times 2^0), \sin(m \times 2^1), \cos(m \times 2^1), \dots, \sin(m \times 2^M), \cos(m \times 2^M)] \quad (2-3)$$

其中 M 为最大编码维度，设定为 5，此值经过多次实验验证后确定，对不同场景均具备高鲁棒性。

随后将强度 I 和位置编码 Pos 进行组合，得到组合输入 $S = I \oplus Pos \in \mathbb{R}^{(2M+1)^I}$ ，确保同时涵盖强度和位置信息。最后，该组合输入被送入多阶段 Transformer 中，每个位置获得独特编码，从而充分捕捉数据顺序的重要特性。

2.3.4 分层表示

HST 模型采用基于像素的 Transformer 处理输入强度。与传统视觉 Transformer 处理图像块不同，该方法将光谱强度划分为单独的像素标记，实现每个像素独立处理，从而提高灵活性和效率。

这些像素标记经过多个 Transformer 块逐层处理，每个模块均利用多头自注意力机制进行并行计算，从而显著提升运算效率。在更深的阶段引入合并层以减

少标记数量，形成分层表示。合并层将相邻标记的特征组合后经线性变换，将标记数量减半，同时加倍通道深度，进一步提炼信息。所有这些阶段共同产生了分层表示。在第 k 阶段，Transformer 的输出为

$$S^{(k)} = \text{Transformer}\{S^{(k-1)}\} \quad (2-4)$$

在这个公式中， $S^{(k)} \in \mathbb{R}^{(2M+1)l/2^k}$ 是第 k 阶段 Transformer 的输出， $S^{(k-1)} \in \mathbb{R}^{(2M+1)l/2^{(k-1)}}$ 是第 k 阶段 Transformer 的输入。另外， $k = 1, 2, 3, \dots$ 。

在实验中使用多阶段表示将 17 个通道转换为单个通道，特征维度设置为 32。在每个阶段，将 Transformer 块重复 $L = 4$ 次，该参数直接影响了模型的复杂度和计算效率。

2.3.5 注意力机制

每个阶段使用多头注意力来促进标记减少。注意力机制允许模型通过分配不同的权重给输入数据的不同部分，聚焦于最相关的部分。

在 Transformer 中，第 k 阶段 Transformer 的输入 $S^{(k-1)}$ 首先通过三个可学习的权重矩阵 W_Q, W_K ，和 W_V 线性转换，分别获得查询 Q 、键 K 和值 V 。将第 i 位置的查询表示为 Q_i 。键和值遵循相同的规则。第 i 和第 j 位置的权重计算如下：

$$\hat{\alpha}_{i,j} = \text{softmax}\left(\frac{Q_i K_j^T}{h}\right) \quad (2-5)$$

其中 h 是查询和键的维度长度。第 i 位置的注意力计算如下：

$$B_i = \sum_j \hat{\alpha}_{i,j} V_j \quad (2-6)$$

输出是值的加权和，使模型能够聚焦于输入数据的最相关部分。在这个阶段结束时，将注意力 $B = \text{concat}(B_1, B_2, \dots, B_l)$ 连接起来，并进一步通过一个线性层传递匹配特征：

$$S^{(k)} = \text{linear}\{B\} \quad (2-7)$$

在每个阶段中，注意力机制可以重复 L 次以产生深度神经网络。在这种情况下，线性层仅在最后一个注意力机制输出时应用。

2.3.6 实现细节

本次实验中采用均方误差（Mean Square Error, MSE）损失函数评估模型性能，其数学表达式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n (s_i - \hat{s}_i)^2 \quad (2-8)$$

其中, n 表示所有数据点的总数, s_i 为真实光谱值, 而 $\hat{s}_i$ 为模型预测结果。该损失函数计算真实与预测光谱间的平方误差平均值, 对较大误差施以更强惩罚, 从而驱动模型在训练过程中尽量缩小二者的差距, 提升预测精度与结果一致性。

在实际的训练过程中, 神经网络模型在 1000 个训练循环内进行了训练, 每次训练的批处理大小为 100。采用了 Adam 优化器来优化模型, 其初始学习率设置为 0.001, 然后每 100 个训练周期, 学习率就会衰减为原来的 0.5 倍, 直到学习率降低到 1×10^{-5} 。模型在 PyTorch 1.8 的环境下实现, 整个训练过程在 NVIDIA 3090 GPU 上进行, 总共大约需要 3 小时的时间。

2.4 高鲁棒性气体种类和浓度监测

解决长波红外高光谱成像挑战的第二个思路是直接在低质量的数据上运行高鲁棒性的后端任务算法, 据此针对长波红外在气体识别领域的应用, 本研究提出了 cross-talk transformer, 利用注意力机制学习目标气体和气体光谱库之间的深层相关性, 并为 USIRS 开发了一个全面的气体辐射传输模型, 以生成大量气体成像数据, 从而充分训练神经网络。

2.4.1 气体辐射传输模型

在化学气体监测时采用了非制冷快照红外成像光谱仪^[82]对特定场景中的气体进行监控。

该系统由三部分组成: 一是红外光谱成像仪, 用于捕捉场景图像; 二是气体室, 用以存储待监测的化学气体; 三是位于气体室后侧的黑体, 作为气体背景。同时, 配备的环境监测仪实时记录温度和湿度, 确保实验数据的准确性。

在实验中, 背景辐射穿过气体室时, 化学气体在不同波长处发生吸收, 随后成像光谱仪捕捉到经过吸收后的辐射信号。通过对采集到的光谱数据分析, 系统能够识别气体种类并预测其浓度, 从而实现对化学气体的全面监测。

在改进的 USIRS 系统中, 气体检测基于存在和不存在化学气体时红外辐射的差异。具体而言, 考虑两种情形:

当检测路径无气体时, 传感器接收的红外辐射主要来自经大气衰减后的背景热辐射 $L_{bg}(\lambda)$ (其中 λ 表示波长) 和大气路径辐射 $L_p(\lambda)$ 。 $L_{bg}(\lambda)$ 用以描述特定波长下的背景热辐射, 而 $L_p(\lambda)$ 则代表大气路径中的辐射成分。

而在检测路径中存在目标气体时, 传感器接收到的热辐射构成更加复杂, 既包含背景辐射, 也涉及气体云的辐射衰减效应和大气路径辐射。气体云指空气中形成的气体集合, 其辐射衰减体现为在传递过程中因吸收和散射造成能量损失。

通过对比这两种情形下传感器接收的红外辐射, 能够判断目标气体的存在, 从而实现有效检测。这一原理为 USIRS 系统提供了理论依据。

L_{off} 表示在检测路径中无气体存在时传感器接收到的红外辐射总量, 其表

达为

$$\begin{aligned} L_{off}(\lambda) &= L_{bg}(\lambda) \cdot \tau_a(\lambda) + L_p(\lambda) \\ &= B(\lambda, T_{bg}) \cdot \varepsilon_{bg}(\lambda) \cdot \tau_a(\lambda) + L_p(\lambda) \end{aligned} \quad (2-9)$$

其中, L_{bg} 为背景辐射, 代表在无其他干扰条件下环境自发的辐射; $B(\lambda, T_{bg})$ 是温度为 T_{bg} 时黑体的发射辐亮度, 根据普朗克定律结合发射率 ε_{bg} 计算得出。 $\tau_a(\lambda)$ 为大气透过率, 其在 7–14 μm 波段主要受水汽影响, 水汽信息可由环境湿度获得。 $L_p(\lambda)$ 是大气路径辐射, 指通过大气传输引入的附加辐射。

当检测路径中存在目标气体时, 传感器接收的热辐射主要由三部分构成: 背景辐射、气体辐射和大气路径辐射, 其数学表达式为

$$\begin{aligned} L_{on}(\lambda) &= L_{bg}(\lambda) \cdot \tau_{gas}(\lambda) \cdot \tau_a(\lambda) + L_{gas}(\lambda) \cdot \tau_a(\lambda) + L_p(\lambda) \\ &= B(\lambda, T_{bg}) \cdot \varepsilon_{bg}(\lambda) \cdot \tau_{gas}(\lambda) \cdot \tau_a(\lambda) \\ &\quad + B(\lambda, T_{gas}) \cdot (1 - \tau_{gas}(\lambda)) \cdot \tau_a(\lambda) \\ &\quad + L_p(\lambda) \end{aligned} \quad (2-10)$$

其中:

$L_{bg}(\lambda)$ 是背景辐射, 表示即使存在气体, 环境自发的背景热辐射仍会被传感器捕获。 $L_{gas}(\lambda)$ 是气体辐射, 描述目标气体自身辐射的热能, 其部分贡献受到气体透射率 $\tau_{gas}(\lambda)$ 的调制。 $L_p(\lambda)$ 是大气路径辐射, 指通过大气传输引入的附加辐射。

在存在目标气体的检测路径中, 传感器接收到的辐射 $L_{on}(\lambda)$ 可分为三部分: 首先, 经气体和大气双重衰减后的背景辐射 L_{bg} ; 其次, 经相应衰减后气体自身产生的辐射 L_{gas} ; 最后, 大气路径辐射 $L_p(\lambda)$ 。其中, 假设气体的发射率等于其吸收率, 即气体发射率为 $1 - \tau_{gas}$, τ_{gas} 表示气体透射率, 反映气体对辐射的吸收能力, 其大小决定了气体发射率的反比变化。

大气路径辐射 $L_p(\lambda)$ 涉及大气成分、温度、湿度、压力等多种因素, 既未知又难以精确估计。为规避直接确定 $L_p(\lambda)$ 的困难, 本文采取了从无气体状态下的辐射与有气体状态下的辐射之差入手的策略, 从而利用已知信息推导出目标气体引起的辐射变化。

$$\begin{aligned} \Delta L(\lambda) &= L_{off}(\lambda) - L_{on}(\lambda) \\ &= [B(\lambda, T_{bg}) \cdot \varepsilon_{bg}(\lambda) - B(\lambda, T_{gas})] \cdot \tau_a(\lambda) \cdot [1 - \tau_{gas}(\lambda)] \end{aligned} \quad (2-11)$$

在实际检测中, 传感器接收到的辐射并非全部转化为探测器的信号, 主要原因在于光学元件对入射辐射存在固有的衰减效应。这些元件无法完全透过所有

辐射，部分能量会被吸收或反射，导致实际接收信号弱于原始辐射。此外，红外探测器本质上为热元件，其工作原理是将接收到的热辐射转换为电信号，转换过程依赖于热电效应。当探测器吸收辐射能量后，部分热能转化为电能，从而形成可测量的输出。基于上述过程，描述红外探测器辐射能到电能转换的关系公式为：

$$DN_{off}(\lambda_i) = G_{gain}(\lambda_i) \cdot \int L_{off}(\lambda) \cdot \tau_s(\lambda_i, \lambda) d\lambda B_{bias}(\lambda_i) + \mathcal{N}_0(\lambda_i) \quad (2-12)$$

$$DN_{on}(\lambda_i) = G_{gain}(\lambda_i) \cdot \int L_{on}(\lambda) \cdot \tau_s(\lambda_i, \lambda) d\lambda B_{bias}(\lambda_i) + \mathcal{N}_1(\lambda_i) \quad (2-13)$$

G_{gain} 是通道 i 的增益，用于信号放大。 B_{bias} 是通道 i 的偏置，能够调整信号的基线。 τ_s 是传感器中所有光学元件的透过率。 λ_i 是通道 i 的中心波长。 $\mathcal{N}_0(\lambda_i)$ 和 $\mathcal{N}_1(\lambda_i)$ 是表示噪声的随机变量，本文假设噪声服从高斯随机分布。

通过辐射校准可以建立探测器像素接收到的辐亮度与其响应之间的关系。为此需要引入了辐射校准系数 a 和 b ，它们直接反映了两者之间的定量联系。具体操作可参照 Yang^[82] 提供的辐射校准方法。具体公式描述了辐射校准系数 a 和 b 与探测器像素辐亮度及响应之间的对应关系：

$$L_{on/off}(\lambda_i) = a(\lambda_i) \cdot DN_{on/off}(\lambda_i) + b(\lambda_i) \quad (2-14)$$

本文中主要采用公式2-11进行具体计算，其中 $\Delta L(\lambda)$ 为设备测得的辐射亮度变化量。背景辐射 $B(\lambda, T_{bg})$ 可利用普朗克定律估算，该定律描述了物体在特定温度下的辐射能量分布；背景发射率 ϵ_{bg} 也可以进行估计。在以上参数已知的情况下，就能通过计算直接求得气体吸收率 τ_{gas} 。

根据比尔-朗伯定律，气体透过率 $\tau_{gas}(\lambda)$ 表示为公式2-15。其中， $\alpha(\lambda)$ 为气体吸收系数，反映了分子在特定波长下的吸收能力； C 为气体浓度（与压力密切相关）； L 为光学长度，即辐射能在气体中的传播距离。该透过率直接反映气体对辐射能的吸收情况。

$$\tau_{gas}(\lambda) = e^{-\alpha(\lambda)CL} \quad (2-15)$$

为了简化计算，通常用泰勒展开将上述公式线性化，但该近似仅适用于微量气体泄漏。对于大规模泄漏，由于气体浓度和压力变化呈非线性关系，线性近似会产生较大误差。为此采用高斯-牛顿算法（见文献^[83]）对气体浓度进行非线性优化计算，以提高精度和准确性。

此外，光学元件透过率 τ_s 也在数据模型中起关键作用。传感器光圈辐射经过平窗、望远镜透镜、准直透镜及透镜阵列等组件时发生衰减，随后通过滤波器

阵列将不同波长分离，再由探测器阵列获取精确信息。因此， τ_s 表示辐射在经过各光学元件后的衰减程度，其计算公式如下：

$$\tau_s(\lambda_i, \lambda) = \tau_w(\lambda_i, \lambda) \cdot \tau_t(\lambda_i, \lambda) \cdot \tau_c(\lambda_i, \lambda) \cdot \tau_l(\lambda_i, \lambda) \cdot \tau_f(\lambda_i, \lambda) \quad (2-16)$$

在模型中， $\tau_w(\lambda_i, \lambda)$ 、 $\tau_t(\lambda_i, \lambda)$ 、 $\tau_c(\lambda_i, \lambda)$ 、 $\tau_l(\lambda_i, \lambda)$ 和 $\tau_f(\lambda_i, \lambda)$ 分别表示平窗、望远镜透镜、准直透镜、透镜阵列和滤波器阵列的透过率，这些参数均由实验室中独立测量得出，并随辐射波长变化而变化，其中 λ_i 为初始波长， λ 为经过光学元件后的波长。

2.4.2 神经网络结构

现代机器学习中，基于注意力机制的 Transformer 已被证明能深度理解复杂数据关系^[52]。基于此，本文提出了 Class-talk Transformer 方法，用于深入学习气体光谱库信息与输入强度间的复杂关系，实现高精度气体种类识别与浓度检测。

为图2-4中展示了 Class-talk Transformer 的整体结构和组成部分。Class-talk Transformer 主要由三部分组成。首先是卷积神经网络 (Convolutional Neural Networks, CNN) 特征提取器，利用卷积神经网络提取输入数据的关键特征。其次是 Transformer 编码器，对 CNN 提取的特征进行进一步编码与转换。最后是多层感知机 (Multi-Layer Perception, MLP) 解码器，将编码信息解码为气体种类和浓度数据。

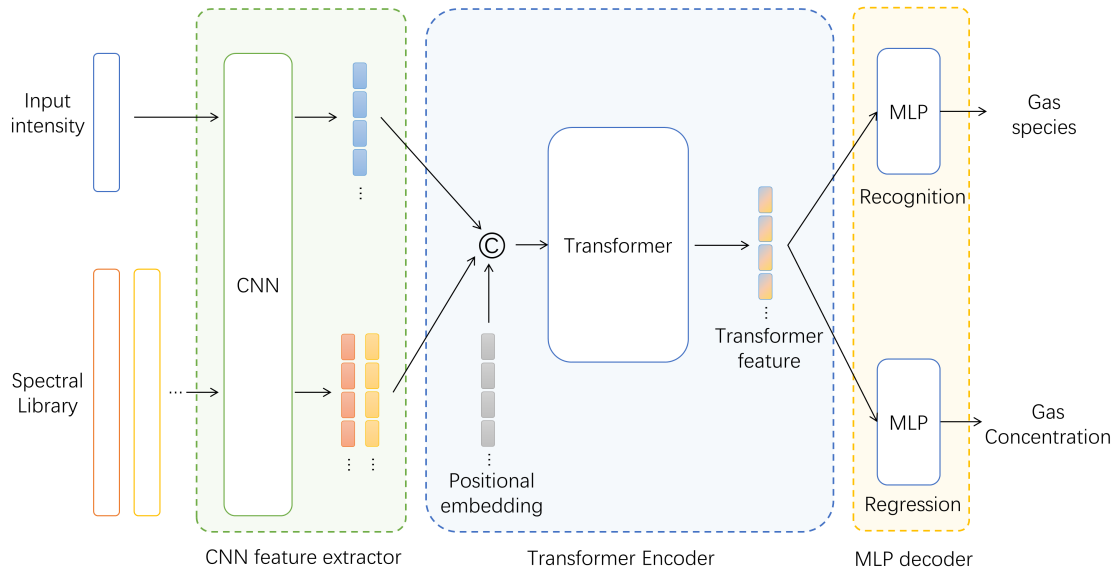


图 2-4 Class-talk Transformer 结构示意图

Figure 2-4 Diagram of Class-talk Transformer Structure

基于在2.4.1中提出的物理成像模型，仪器在接收和处理信号时会受到多种噪声干扰。因此，首先使用浅层卷积神经网络从受噪声影响的信号中提取浅层特

征。提取完成后，Transformer 编码器将信号特征与光谱库特征融合，学习它们之间的深层相关性，生成新的 Transformer 特征。随后，两种多层感知机模型分别利用这些 Transformer 特征完成气体种类的分类和浓度的回归预测。

在 CNN 特征提取器中，为克服噪声采用四个卷积块，每个卷积块结构相同，包括 3×3 的膨胀卷积层，用于提取局部特征；批量归一化层，降低内部协变量偏移；ReLU 激活层，增强模型非线性表达能力。这四个卷积块的通道数依次为 16、32、64 和 64，从低信噪比信号中提取出尺寸为 64×3 的浅层特征，经展平后得到 192 维特征（记为 F ）。同时，光谱库包含 24 种气体光谱，通过同一 CNN 提取器处理后获得尺寸为 192×24 的浅层特征。这些浅层特征为后续 Transformer 编码器的深层特征学习提供了基础。

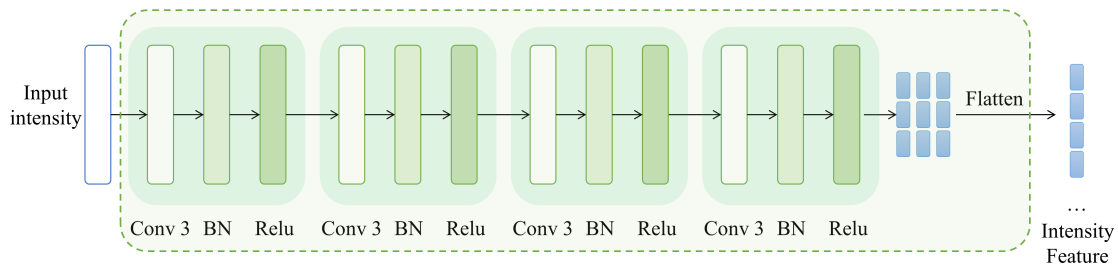


图 2-5 Transformer 特征提取器的结构示意图

Figure 2-5 Schematic Diagram of the Transformer Feature Extractor Structure

在 Class-talk Transformer 中，Transformer 编码器是关键组件，负责捕捉输入信号与光谱库间的复杂关系。为了提升模型表达能力与序列理解，引入位置嵌入技术。位置嵌入通过三角变换将每个特征的位置（记为 k ）转化为新的特征，帮助模型捕捉数据序列的顺序和周期性信息。借助位置嵌入，Transformer 不仅能学习信号与光谱间的复杂关联，还能准确捕捉其时序特性，从而提升整体任务的表现。

$$\begin{aligned}
 pos(k) = & [\sin(k \times 2^0), \cos(k \times 2^0), \\
 & \sin(k \times 2^1), \cos(k \times 2^1), \\
 & \dots \\
 & \sin(k \times 2^N), \cos(k \times 2^N)]
 \end{aligned} \tag{2-17}$$

在 Class-talk Transformer 模型中，参数 N 表示最大编码维度，经过大量实验验证，将 N 固定为 5，这一设置在确保模型计算效率的同时，也能有效捕捉输入信号与光谱库间的复杂关系。

模型首先将 CNN 提取的强度特征与光谱库特征通过位置嵌入后融合为联合特征表示，然后利用注意力机制学习特征间关联。注意力机制使用查询矩阵 Q 、键矩阵 K 和值矩阵 V ，其中 Q 用于匹配 K 以筛选关键信息， V 则反映匹配强度。为提升性能，采用多头注意力策略，使模型从多角度捕捉关联，尽管本文中

仅讨论第一个注意力头。所有的 Q 、 K 、 V 均为可训练参数，能在训练过程中自动学习。

具体实现中用如下公式计算第 i 个特征与第 j 个特征之间的匹配系数，该系数反映了它们的关联程度：

$$\hat{\alpha}_{i,j} = \frac{\exp(Q^i K^j)}{\sum_j (Q^i K^j)} \quad (2-18)$$

为深入理解注意力机制，此处引入匹配特征 b_i ，它反映第 i 个特征与所有其他特征之间的关联程度。具体来说， b_i 是将与第 i 个特征相关的所有匹配系数与对应值矩阵 V 的加权和。计算过程为：首先计算每个特征与第 i 个特征的匹配系数，再将这些系数与 V 相乘，最后对乘积求和即得 b_i 。匹配特征 b_i 帮助模型捕捉特征间的关联，从而更好地理解输入信号与光谱库间的复杂关系，这是提升模型性能的重要因素。匹配特征 b_i 的计算公式如下：

$$b_i = \sum_j \hat{\alpha}_{i,j} V_j \quad (2-19)$$

在 Class-talk Transformer 中，通过全连接网络对匹配特征进行多次传递，以增强网络的学习能力。该过程重复 L 次，实现深层匹配特征提取，使模型能从多层次、多角度理解输入信号与光谱库间的复杂关系，并最终输出 Transformer 特征表示。具体实验设置如下：Transformer 嵌入维度设置为 192，确保充分捕捉信号与光谱间的关系；注意力头数量设置为 8，以多角度捕捉特征关联；Transformer 编码器重复次数设置为 2，实现深层特征提取；全连接网络隐藏层神经元数量设置为 768，提升匹配特征的表达能力。

最终，Transformer 特征通过两个多层感知器网络分别实现气体识别和浓度预测。气体识别采用一个隐藏层（128 个神经元），经过批量归一化和 ReLU 激活，再通过对数 Softmax 函数输出每种气体的预测概率；浓度预测则由一个由 128、64、32、16 个神经元构成的全连接网络实现，最终输出预测气体浓度。

2.4.3 实现细节

2.4.4 神经网络实现

采用 PyTorch 实现 Class-talk Transformer 模型，并在模拟数据上训练以预测化学气体类型和浓度。模拟数据有助于捕捉气体的复杂性质，从而提升模型性能。训练设置如下：批处理大小设置为 1000 样本，旨在平衡训练速度与性能；网络初始化为随机状态，并使用 ADAM 优化器（学习率 1×10^{-4} ，逐 epoch 乘以 0.9， $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ， $\epsilon = 1 \times 10^{-8}$ 。）训练，使模型快速收敛；训练在单块 NVIDIA 1050Ti GPU 上进行，约 10 分钟内收敛。

训练过程中，记录训练、验证和测试集的损失（详见图2-6），损失均迅速下降，并在训练末期达到一致水平，表明模型具有良好的训练效果和泛化能力。

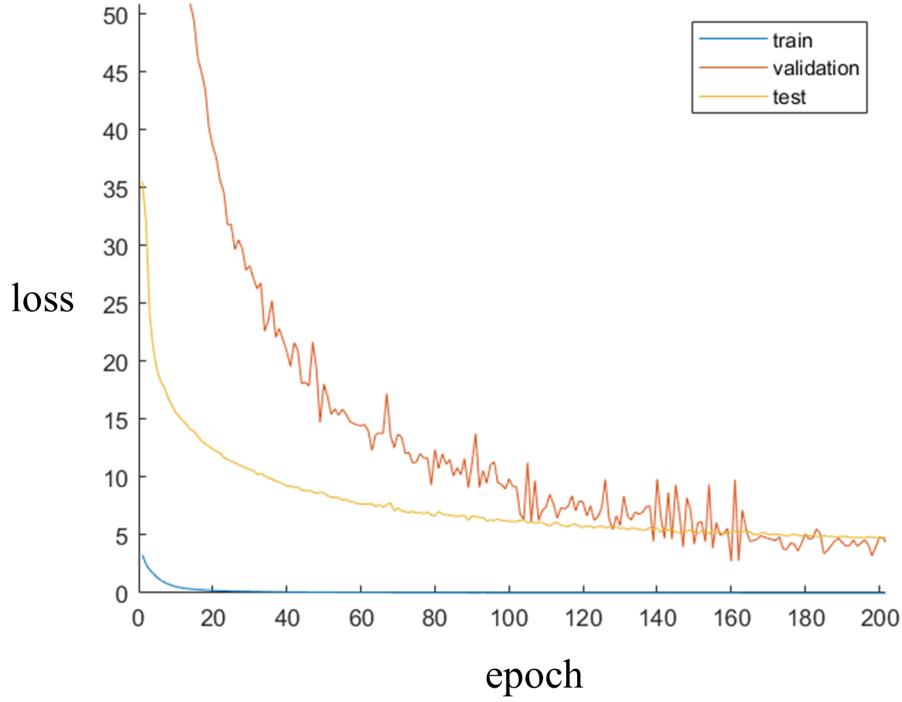


图 2-6 训练、验证和测试过程中的损失曲线

Figure 2-6 Loss Curve during Training, Validation, and Testing Process

2.5 图像重建实验结果

2.5.1 评估指标

为定量评估 Class-talk Transformer 的光谱重建性能，采用常见的相关系数指标。相关系数反映两组光谱间的线性关系，其值范围为 $[-1, 1]$ ，值越接近 1 表示正线性关系越强，接近 -1 表示负线性关系，接近 0 则说明相关性较弱。具体计算公式如下：

$$r = \sum_{i=1}^l \frac{(s_i - \bar{s})(g_i - \bar{g})}{l\sigma_s\sigma_g} \quad (2-20)$$

其中， s 和 g 分别表示目标光谱和真实光谱； s_i 和 g_i 为各索引位置的值； $\bar{s}$ 和 $\bar{g}$ 为平均值； σ_s 和 σ_g 为标准差； l 为光谱长度。通过计算相关系数，能够定量分析模型的光谱重建效果。

除了相关系数，还采用均方根误差 (Root Mean Square Error, RMSE) 评估模型光谱重建的准确性。RMSE 衡量重建光谱与真实光谱之间差异的大小，其定义为差异平方和的平均值再取平方根。较低的 RMSE 值表明模型性能较好，重建结果更接近真实光谱。

$$RMSE = \sqrt{\frac{1}{l} \sum_{i=1}^l (s_i - g_i)^2} \quad (2-21)$$

其中, s 和 g 分别为目标光谱和真实光谱; s_i 和 g_i 表示对应索引处的值; l 为光谱向量长度。通过 RMSE, 可以定量地评估模型的光谱重建效果, 从而进一步优化模型性能。

峰值信噪比 (Peak Signal Noise Ratio, PSNR) 是评估光谱重建模型的一项重要指标。其衡量了真实光谱与重建光谱之间的误差, 通过比较原始数据最大可能功率与噪声功率 (通常通过均方误差表示) 的比值得出, 结果以分贝 (dB) 表示。PSNR 值越高, 说明重建光谱质量越接近真实光谱。

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_s}{\sqrt{MSE}} \right) \quad (2-22)$$

其中, MAX_s 代表光谱数据可能的最大值。对于归一化到 $[0, 1]$ 范围内的光谱数据, MAX_s 取 1。通过 PSNR 能够直观地评价模型在光谱重建上的表现。

均方误差 (Mean Square Error, MSE) 用于量化目标光谱与真实光谱在各采样点上差异的平均值, 其计算步骤为: 对每一采样点求误差平方, 再对所有误差平方求平均。公式为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (s_i - g_i)^2 \quad (2-23)$$

其中, n 为光谱长度 (即波长点数), s_i 和 g_i 分别表示第 i 个波长点上的重建光谱和真实光谱的强度值。将 MSE 与最大光谱强度值 MAX_s (对于归一化数据, MAX_s 通常为 1) 结合, 可通过 PSNR 公式评价光谱重建质量。

PSNR 值越高, 说明噪声引入的失真越低, 重建光谱质量越佳。但需要注意, 即便 PSNR 较高, 也可能存在整体亮度差异等问题, 因此应结合其他指标和直观观察进行综合评估。

2.5.2 仿真数据结果

采用仿真数据评估方法有效性。数据集分为四部分: 第一部分是从 ASTER 长波红外光谱库中选取的 1289 个样本, 这些样本具有独特红外特征。ASTER 库整合了 2300 多种自然和人造材料的光谱, 波长范围为 $0.4\text{--}15.4\ \mu\text{m}$, 涵盖岩石、矿物、土壤、陨石、植被、雪冰等多种物质。第二部分是 1000 个符合普朗克定律的黑体辐射样本, 温度范围为 100K 至 1000K。第三部分是 24 种长波红外气体的发射光谱样本, 涉及甲烷 (CH_4)、乙烯 (C_2H_4)、氨气 (NH_3)、六氟化硫 (SF_6)、环己烷 (C_6H_{12})、甲醇 (CH_3OH)、丙酮 (CH_3COCH_3)、三甲胺 ($(CH_3)_3N$)、环丙烷 (C_3H_6)、丙烯 (C_3H_6)、反-2-丁烯 (C_4H_8)、丁二烯 (C_4H_6)、环氧乙烷 (C_2H_4O)、二甲胺 ($HN(CH_3)_2$)、氯乙烯 (C_2H_3Cl)、1-丁烯 (C_4H_8)、乙炔 (C_2H_2)、丙炔 (C_3H_4)、二甲醚 (C_2H_6O)、乙醛 (CH_3CHO)、氯乙烷 (C_2H_5Cl)、氯甲烷 (CH_3Cl)、甲胺 (CH_3NH_2) 和二氧化硫 (SO_2)。

等气体。第四部分是基于文献^[29]模型模拟黑体前气体辐射光谱的 2400 个样本。这四部分数据资源丰富，有力支持了方法的评估。

对提出的 HST 进行定性评估，检验其在矿物、黑体、气体发射和气体传输四种光谱数据上的高分辨率重建能力。图 2-7 显示，重建数据与真实数据高度一致，充分证明其有效性与可靠性，并凸显了准确光谱重建在遥感、光谱学和材料识别中的关键作用。

此外，HST 不仅适用于这四种光谱类型，还能处理更多种类的数据，生成高分辨率重建。虽然本文因篇幅限制仅展示四种类型的结果，但后续实验进一步验证了其广泛适用性。

对提出的方法进行严谨定量评估，以验证其优越性，并与各种方法进行了详细比较。图 2-8 展示了 HST 与其他神经网络方法（MLP、CNN、Transformer）及插值方法在重建高岭土、770K 黑体、氯乙烯气体发射和丙烯气体传输四种材料光谱时的表现，同时图例中给出了重建光谱与真实光谱的相关系数，反映了两者的相似程度。

评估结果表明，HST 在所有材料上均实现了最高重建质量，明显优于其他方法。线性和三次插值方法只能提供简单近似，而神经网络的基准方法虽有改进，仍未达到 HST 的性能。这充分证明了 HST 在光谱重建领域的领先地位，也验证了研究方向的正确性。

接下来，在不同噪声水平（0、0.1、0.2、0.5）下，对 HST 及其他方法（线性插值、立方插值、MLP、CNN、Transformer）在仿真数据上的重建性能进行了对比，评价指标包括 RMSE、相关性和 PSNR。表 2-1 显示，HST 在几乎所有噪声水平下均达到最低 RMSE 和最高相关性及 PSNR，表明其重建误差最小、与真实光谱最为接近且信噪比最佳。虽然 CNN 和 Transformer 表现也颇为优秀，但均略逊于 HST。总体而言，HST 在各项指标上均表现出色，再次证明了其在光谱重建领域的优越性。

2.5.3 实验室测试数据结果

在仿真数据测试之后，进一步用实验数据评估方法。实验数据包括：(1) 实验室数据：甲烷（ CH_4 ）、乙烯（ C_2H_4 ）、氨（ NH_3 ）、六氟化硫（ SF_6 ）、三甲胺（ $(\text{CH}_3)_3\text{N}$ ）、环丙烷（ C_3H_6 ）、丙烯（ C_3H_6 ）、反-2-丁烯（ C_4H_8 ）、丁二烯（ C_4H_6 ）、环氧乙烷（ $\text{C}_2\text{H}_4\text{O}$ ）和氯乙烯（ $\text{C}_2\text{H}_3\text{Cl}$ ）这 11 种气体。(2) 现场数据：氨气（ NH_3 ）和六氟化硫（ SF_6 ）这两种气体以及黑体。

实验室中测量数据时以黑体（背景温度 50°C ）作为理想背景，在黑体与原型间设有 0.5 米光学路径的定制气体室。将气体以 0.5%、1% 和 2% 纯度充入气室（对应路径积分浓度分别为 2500、5000 和 10000 ppm-m），在室内温度 23.1°C 、湿度 40% 及 1 大气压的条件下进行测量。

在实验室环境中比较了 HST 与其他方法对气体传输光谱的重建效果（图 2-9）。HST 能从由线性插值生成的低分辨率输入数据中，准确重建环丙烷、丁二

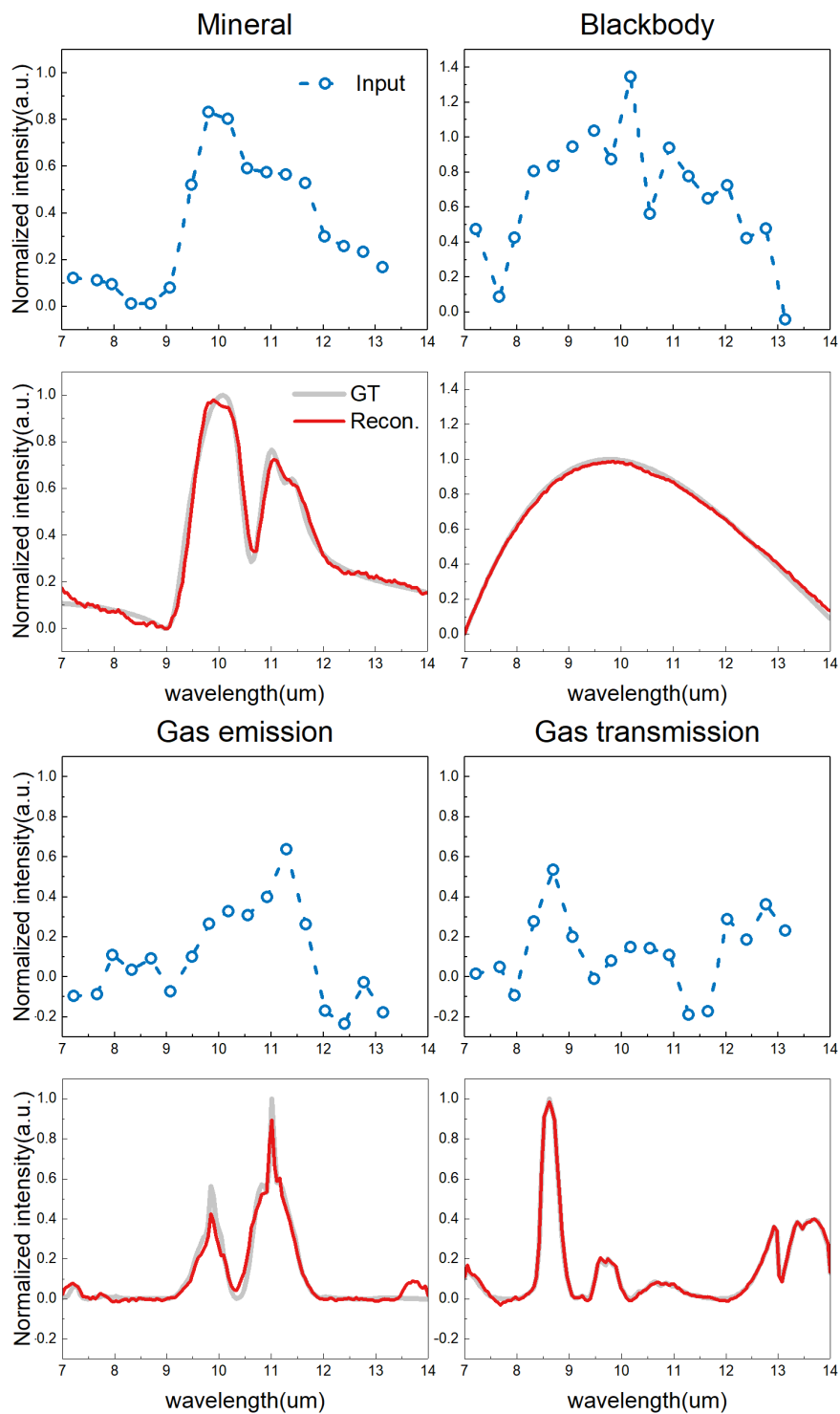


图 2-7 HST 重建四种不同类型的光谱定性分析

Figure 2-7 Qualitative Analysis of Four Different Types of Spectra Reconstructed by HST

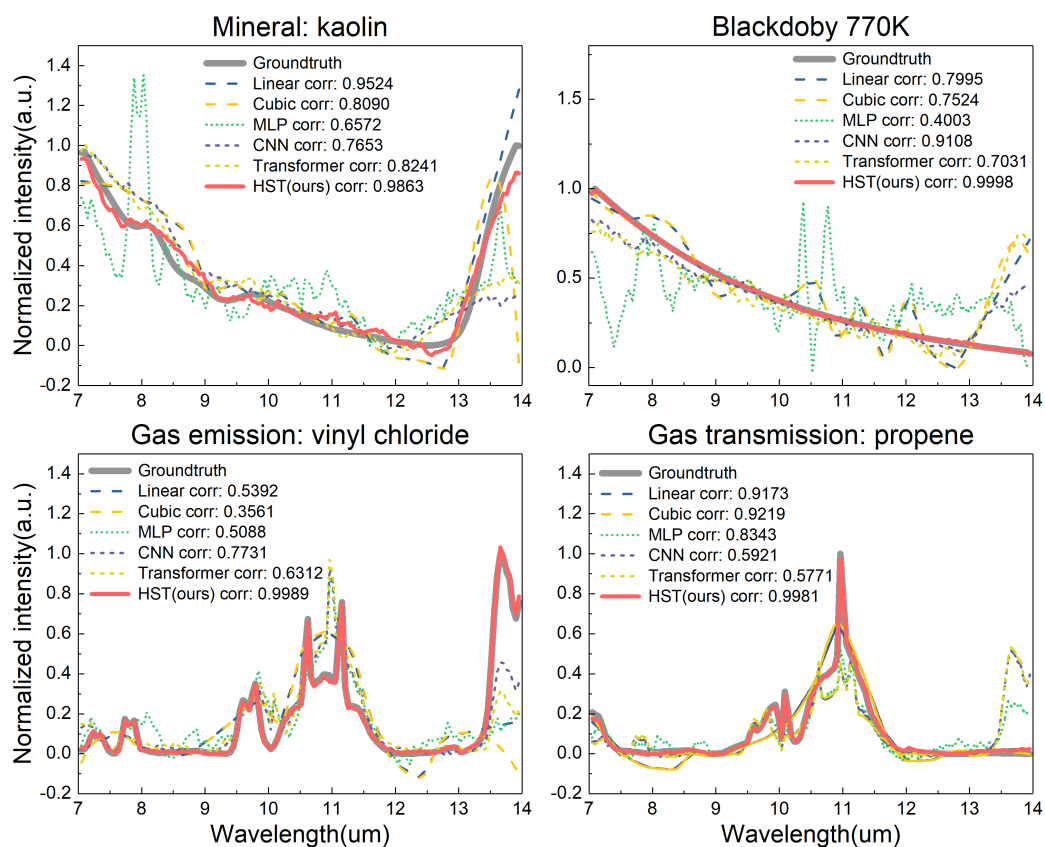


图 2-8 使用仿真数据集比较不同方法

Figure 2-8 Comparing Different Methods Using a Simulated Dataset

表 2-1 在不同噪声水平下使用仿真数据集比较不同方法的表现。
Table 2-1 Comparing the Performance of Different Methods Using a Simulated Dataset at Various Noise Levels.

噪声水平		0	0.1	0.2	0.5
线性插值	RMSE	0.1133	0.1346	0.1793	0.3532
	Correlation	0.9558	0.9433	0.9134	0.8008
	PSNR	17.54	16.28	13.98	8.16
三次插值	RMSE	0.1139	0.1686	0.2629	0.5953
	Correlation	0.9593	0.9180	0.8420	0.6868
	PSNR	17.93	14.33	10.13	2.75
MLP	RMSE	0.1208	0.1130	0.1767	0.2153
	Correlation	0.9568	0.9599	0.9011	0.8420
	PSNR	18.04	18.35	14.63	12.83
CNN	RMSE	0.0103	0.0233	0.0428	0.1076
	Correlation	0.9989	0.9968	0.9906	0.9565
	PSNR	33.87	29.22	24.53	18.03
Transformer	RMSE	0.0061	0.0238	0.0436	0.1089
	Correlation	0.9994	0.9970	0.9904	0.9563
	PSNR	36.51	29.46	24.45	17.99
HST (ours)	RMSE	0.0059	0.0212	0.0378	0.1074
	Correlation	0.9995	0.9976	0.9929	0.9566
	PSNR	37.16	30.41	25.74	18.02

注: 红色: 最佳表现. 蓝色: 次佳表现.

烯、氯乙烯和甲烷这四种气体在 7–14 μm 范围内的高分辨率光谱。与 MLP、CNN 及 Transformer 相比，HST 能有效捕捉细节，例如能清晰显示环丙烷在 9.5 μm 处的峰值，而其他方法未能呈现这一信息。总体而言，HST 在高分辨率光谱重建上表现优异，突显了其在光谱重建领域的应用价值。

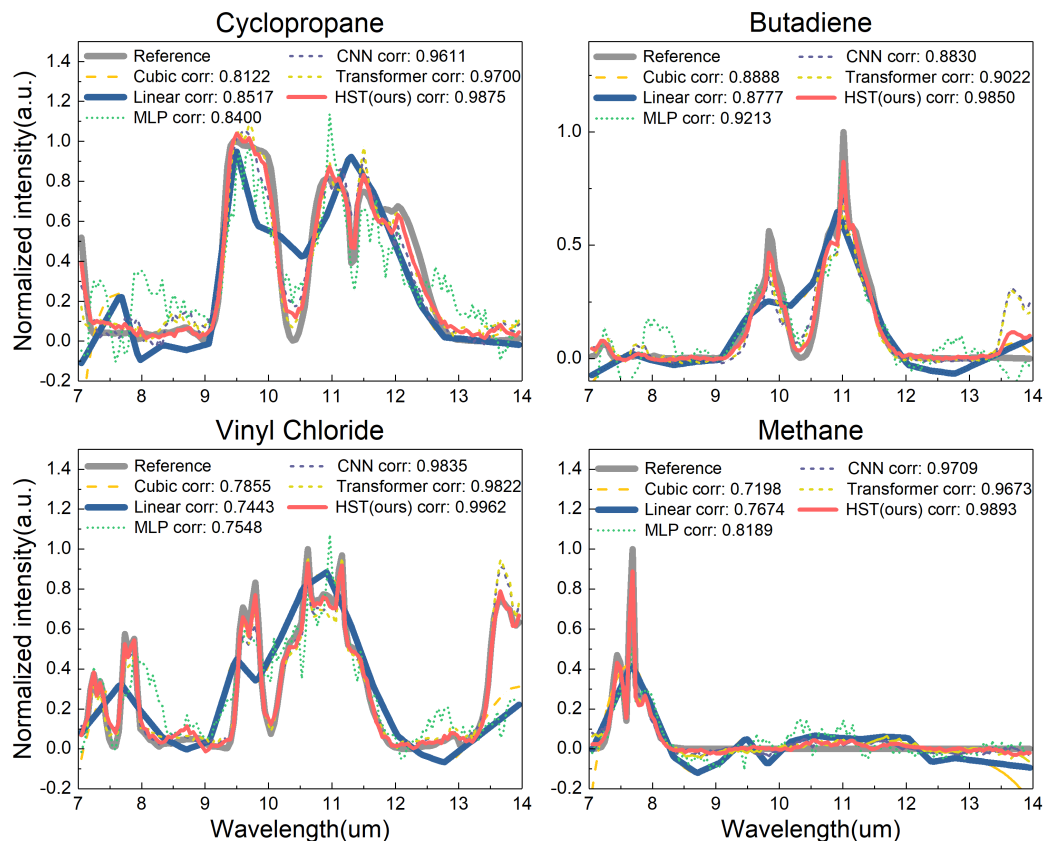


图 2-9 通过实验室数据比较四种不同气体传输光谱的不同方法

Figure 2-9 Comparison of different methods using experimental laboratory data of four different gas transmission spectra

对各种插值和学习方法在 RMSE、相关性和 PSNR 三个指标上的性能进行了定量评估（见表2-2）。结果显示：线性插值和立方插值表现中等，RMSE 分别为 0.1149 和 0.1338，表明误差控制一般。MLP 方法略有改进，其 RMSE 为 0.1241，相关性达到 0.9248，显示了准确性的提升。CNN 方法显著提高了性能，RMSE 降至 0.0437，相关性提升到 0.9866。Transformer 方法也表现不错，RMSE 为 0.0422，相关性为 0.9880。提出的 HST 方法表现最佳，RMSE 仅为 0.0333，相关性最高为 0.9915，PSNR 达 26.78，证明在误差最小、相似度最高和信噪比最优方面均领先于其他方法。该表格中的结果充分证明了 HST 在高分辨率光谱重建中的有效性和优越性。

2.5.4 外场测试数据结果

图2-10展示了使用现场数据对光谱成像在复杂场景中的测试过程。实验在距离现场约 100 米、离地 15 米处进行，拍摄时地温约 35°C、气体温度 22°C，强烈

表 2-2 使用实验室数据定量比较不同方法。

Table 2-2 Quantitatively Comparing Different Methods Using Laboratory Data.

方法	指标	性能
线性插值	RMSE	0.1149
	Correlation	0.9386
	PSNR	18.30
三次插值	RMSE	0.1338
	Correlation	0.9059
	PSNR	16.28
MLP	RMSE	0.1241
	Correlation	0.9248
	PSNR	17.67
CNN	RMSE	0.0437
	Correlation	0.9866
	PSNR	24.87
Transformer	RMSE	0.0422
	Correlation	0.9880
	PSNR	25.32
HST (ours)	RMSE	0.0333
	Correlation	0.9915
	PSNR	26.78

注: 红色: 最佳表现. 蓝色: 次佳表现.

的热对比有利于突出气体目标，但也引入了背景和泊松噪声。现场释放了 40,000 ppm 的 SF_6 和 NH_3 （分别用红、绿框标记）。

利用仿真数据训练后的 HST 模型，从每个像素的噪声输入中重建高分辨率光谱，并针对道路、 SF_6 和 NH_3 三个目标进行重建。图中显示，尽管输入光谱因噪声和失真与参考光谱存在差异，重建结果仍与参考光谱高度一致，定量结果进一步证明了该方法在保持光谱特征和准确重建方面的优越性。

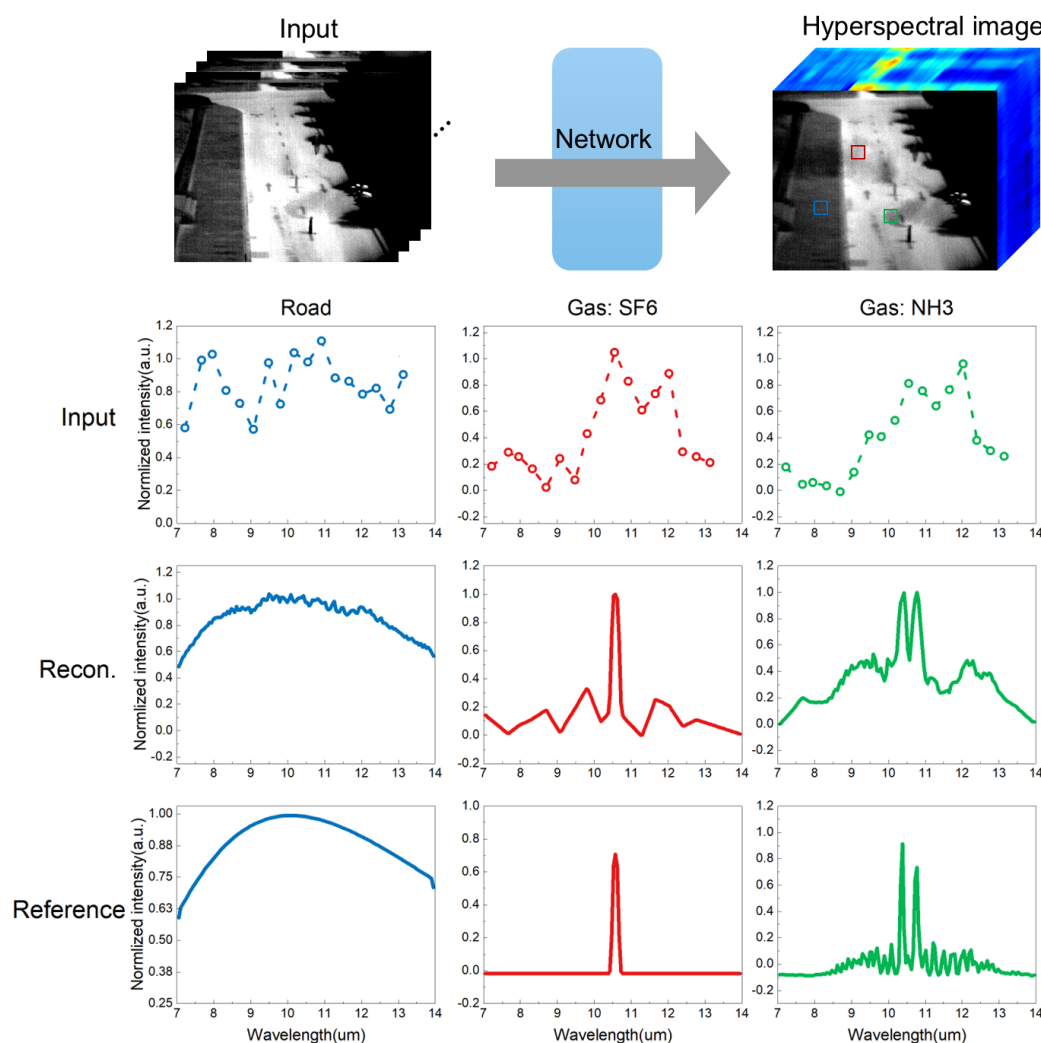


图 2-10 外场实验展示 HST 的应用。

Figure 2-10 Illustration of HST using field experiments.

2.6 气体监测实验结果

2.6.1 实验装置

图 2-11(a) 展示了传感器的详细结构，由两台不同光谱带的 USIRS 并行安装构成，用于获取更丰富的气体光谱数据。该系统主要包括伸缩镜头、准直镜头、滤光片阵列、镜头阵列和非制冷 VOx FPA 探测器，共同置于充满氮气、恒温控制的环境中。工作过程中，场景的红外辐射先经伸缩镜头聚焦至主像平面，再由

准直镜头准直，随后通过滤光片阵列过滤和镜头阵列汇聚，最终在探测器上成像。系统波长覆盖 7–14 μm ，大多数气体在此范围内展现独特吸收峰，其光谱带包括 17 个 700 nm 带通通道（中心波长分别为 7.22、7.67、7.96、8.33、8.70、9.07、9.48、9.81、10.18、10.55、10.92、11.29、11.66、12.03、12.40、12.77 和 13.14 μm ）以及一个覆盖全波段的宽带通道。探测器阵列尺寸为 1024×1280 像素，经剔除遮挡部分后单通道空间分辨率为 384×307 像素。在 25°C 环境下，带通光谱通道的灵敏度超过 220 mK，宽带通道超过 58 mK，平均响应时间低于 125 ms，确保了高速、高效的数据采集。

为精确反演气体浓度，设计制造了特制气室（图 2-11(c)）。气室采用硬铝外壳，配置进出气阀、压力与温度传感器、磁旋风分离器、红外和可见窗口。内部安装有温度和压力传感器实现实时监测。气室光学路径为 0.6 米，两端红外窗口平均透过率达 90%。为确保不同密度气体均匀分布，气室内设有磁旋风分离器，通过高速旋转气流分离气体。

实验中，选用了面积 300×300 mm^2 的黑体作为背景，其工作温度范围 0–180°C、温度分辨率达 0.001°C，保障了较大温度范围内的高精度测量。同时，在气室旁配置环境监测器，用于实时监控温度和湿度，确保实验环境的稳定性和数据准确性。

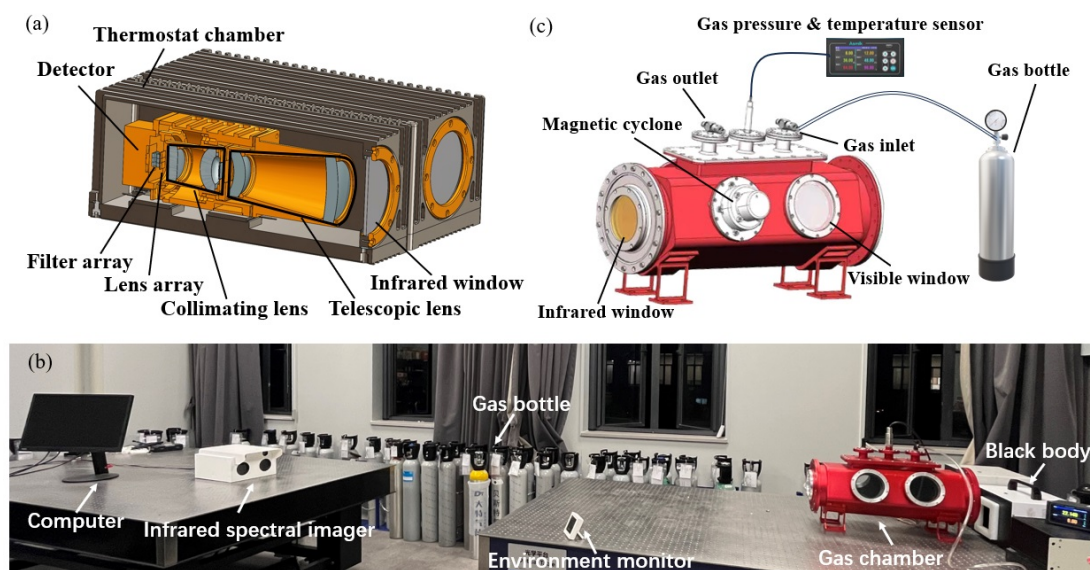


图 2-11 非制冷快照红外光谱仪气体实验装置

Figure 2-11 Gas Experiment Device for Non-Refrigerated Snapshot Infrared Spectrometer

2.6.2 实验数据

为全面评估所提出的化学气体识别与浓度预测算法，采用了两种数据集：一是由 USIRS 采集的真实实验数据，二是基于第2.4.1节中的辐射传输模型生成的模拟数据。

实验数据由杭州百斯特气体有限公司提供，涵盖 11 种适于实验室环境下使

用的气体, 包括甲烷 (CH_4)、乙烯 (C_2H_4)、氨气 (NH_3)、六氟化硫 (SF_6)、三甲胺 ($(\text{CH}_3)_3\text{N}$)、环丙烷 (C_3H_6)、丙烯 (C_3H_6)、反-2-丁烯 (C_4H_8)、丁二烯 (C_4H_6)、环氧乙烷 ($\text{C}_2\text{H}_4\text{O}$) 和氯乙烯 ($\text{C}_2\text{H}_3\text{Cl}$)。浓度范围为 200 至 30,000 ppm, 共计 11,680 个样本。实验设定为: 气体温度 22°C , 背景温度 50°C , 环境温度为 $25\pm 0.3^\circ\text{C}$ 、湿度 40%, 同时需考虑压力传感器存在 0.01 MPa 的误差。

模拟数据集约包含 2,400,000 个样本, 涵盖 24 种气体。模拟条件与实验相同 (气体温度 22°C 、背景温度 50°C), 且气体浓度服从均值为 0、标准差为 1 的正态分布, 范围为 10 至 540,000 ppm, 能覆盖实际应用中的大部分情况。

对比甲烷 (CH_4) 在不同浓度下的模拟数据与测量数据, 可见二者高度一致, 这验证了模拟数据集的有效性和模拟方法的准确性。高质量且大规模的数据为神经网络模型的训练提供了充分支持, 从而有助于更精确地理解和预测化学气体的行为。

2.6.3 评价指标

在评估神经网络性能时, 采用准确率作为评价指标。准确率是分类问题中常用且直观的衡量标准, 其计算公式为:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-24)$$

其中, TP 为正确识别的正类样本数, TN 为正确识别的负类样本数, FP 表示将负类误判为正类的数量, FN 表示将正类误判为负类的数量。借助这四个指标, 能够清晰揭示模型在处理正负样本时的识别效果, 从而更全面地评估其性能。

在浓度预测任务中, 采用均方根误差和平均绝对误差来衡量模型性能。这两项回归指标有效揭示出预测的精度与稳定性, 其计算公式分别为:

$$RMSE = \|c_{pred} - c_{gt}\|_2 \quad (2-25)$$

其中, c_{pred} 表示模型预测的浓度值, 而 c_{gt} 为模拟数据或实验测量得到的真实浓度。这两个指标为全面评估模型表现提供了定量依据。

$$MAE = \|c_{pred} - c_{gt}\|_1 \quad (2-26)$$

为了全面评价所提出的 Cross-talk Transformer 性能, 选用多层感知机和卷积神经网络作为对比模型, 同时采用准确率、均方根误差和平均绝对误差等指标进行评估。

具体而言, MLP 主要由全连接层组成, 适合处理一维数据, 而 CNN 则具备处理二维图像等复杂数据结构的能力。然而, MLP 与 CNN 在处理长序列数据和捕捉复杂模式方面存在局限。相比之下, Cross-talk Transformer 利用 Transformer 结构, 有效解决了上述问题, 实现了对长序列及复杂模式的更佳处理。

2.6.4 仿真数据结果

为全面评估 Cross-talk Transformer 的性能，选用包含 24 种化学气体不同浓度数据的模拟数据集进行实验。该数据集样本数量充足、种类丰富，为模型测试提供了理想条件。

实验结果见图 2-12 和表 2-3。数据显示，Cross-talk Transformer 在所有 24 种气体的识别任务中准确率均超过 98%，平均准确率达到 99.77%。这一结果充分证明了模型在复杂气体识别任务中的优异表现与较强的泛化能力，无论气体种类或浓度均能实现高精度识别。

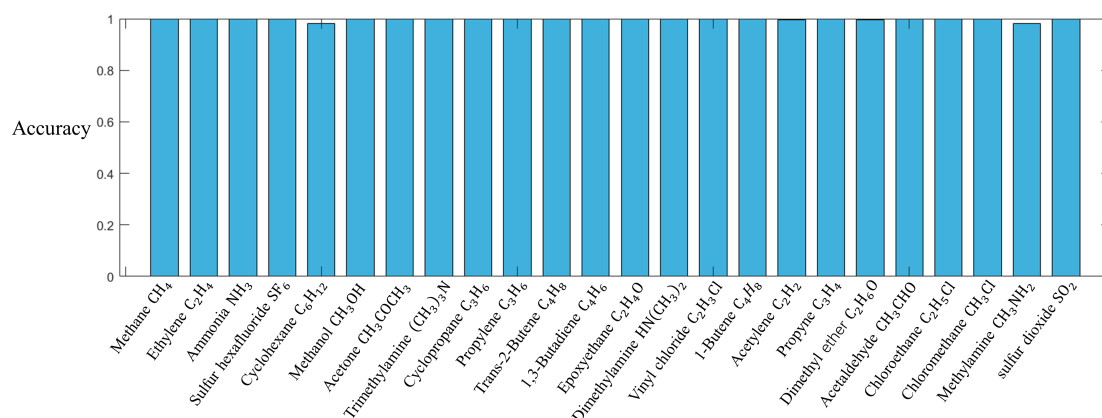


图 2-12 Class-talk Transformer 在仿真数据集上的气体种类识别结果

Figure 2-12 Gas Type Identification Results of Class-Talk Transformer on Simulated Dataset.

图 2-13 展示了模型对甲烷、乙烯、氨气、六氟化硫、三甲胺、环丙烷、丙烯、反-2-丁烯和丁二烯这九种典型化学气体浓度的预测效果。这些气体广泛应用于制造业，浓度范围从 10 ppm 到 25,000 ppm，为模型预测带来了极大挑战。实验结果表明，Cross-talk Transformer 在这些气体的浓度预测中均实现了高精度与稳定性，验证了其在实际环境监测与化学泄露预警中的应用潜力，有助于及时识别泄露风险并保障工厂安全。

2.6.5 实验数据结果

在辐射传输模型模拟数据集上训练完 Cross-talk Transformer 后，继续在实际采集的实验数据集上进行了评估。该数据集涵盖 11 种典型化学气体在不同浓度下的数据，能够更真实地反映模型的应用效果。

如图 2-14 和表 2-3 所示，实验结果显示，在甲烷、氨气、六氟化硫、三甲胺、环丙烷、丙烯、反-2-丁烯和环氧乙烷这八种气体的识别任务中，模型准确率均超过 98%，充分体现了其卓越性能与高度可靠性。而对于乙烯、丁二烯和氯乙烯这三种气体，尽管识别准确率略低，但仍达到 92%，证明了模型在不同气体任务上的通用性和适应性。

图 2-15 展示了 Cross-talk Transformer 在实际测量数据上的浓度预测表现。该数据覆盖 100 ppm 至 30,000 ppm 的广泛浓度范围，涉及甲烷、乙烯、六氟化硫、

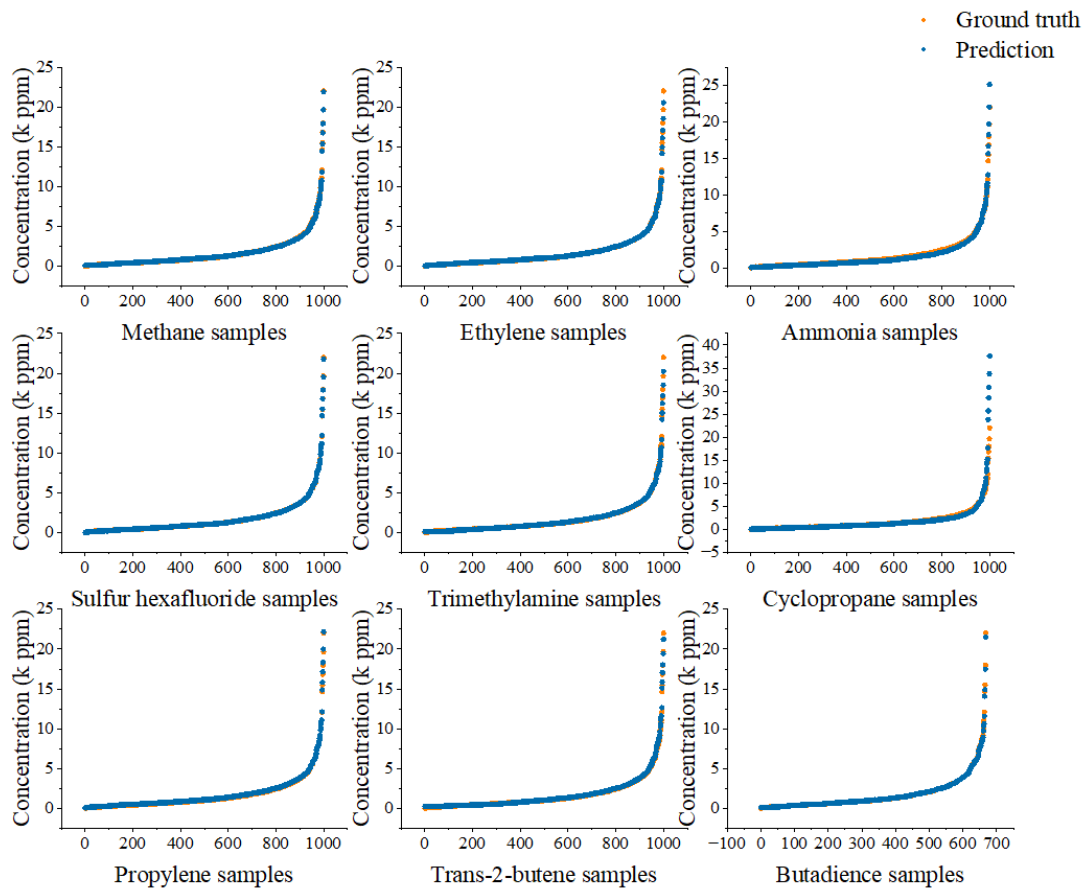


图 2-13 Cross-talk Transformer 在模拟数据集上的浓度预测结果

Figure 2-13 Sample Figure

表 2-3 Class-talk Transformer 在模拟和实验数据集上的识别性能。

Table 2-3 Recognition Performance of Class-Talk Transformer on Both Simulated and Measured Datasets.

准确率	模拟	实验
甲烷 (CH ₄)	100.00%	100.00%
乙烯 (C ₂ H ₄)	100.00%	91.95%
氨气 (NH ₃)	100.00%	100.00%
六氟化硫 (SF ₆)	100.00%	100.00%
环己烷 (C ₆ H ₁₂)	98.24%	-
甲醇 (CH ₃ OH)	99.97%	-
丙酮 (CH ₃ COCH ₃)	99.91%	-
三甲胺 ((CH ₃) ₃ N)	99.98%	100.00%
环丙烷 (C ₃ H ₆)	100.00%	99.47%
丙烯 (C ₃ H ₆)	100.00%	100.00%
反-2-丁烯 (C ₄ H ₈)	99.99%	100.00%
丁二烯 (C ₄ H ₆)	100.00%	98.64%
环氧乙烷 (C ₂ H ₄ O)	99.95%	99.22%
二甲胺 (HN(CH ₃) ₂)	99.86%	-
氯乙烯 (C ₂ H ₃ Cl)	100.00%	93.99%
1-丁烯 (C ₄ H ₈)	99.95%	-
乙炔 (C ₂ H ₂)	99.66%	-
丙炔 (C ₃ H ₄)	99.90%	-
二甲醚 (C ₂ H ₆ O)	99.55%	-
乙醛 (CH ₃ CHO)	99.89%	-
氯乙烷 (C ₂ H ₅ Cl)	99.93%	-
氯甲烷 (CH ₃ Cl)	99.82%	-
甲胺 (CH ₃ NH ₂)	98.24%	-
二氧化硫 (SO ₂)	99.75%	-

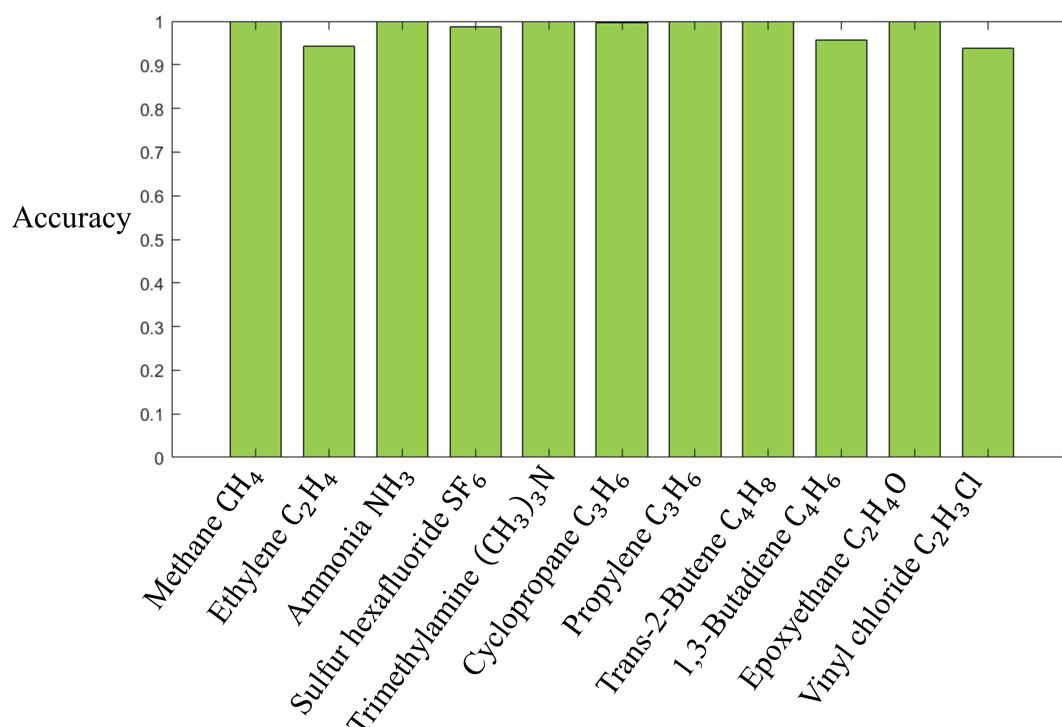


图 2-14 Class-talk Transformer 在实测数据集上的气体种类识别结果

Figure 2-14 Gas Type Identification Results of Class-Talk Transformer on Measured Dataset.

环丙烷、丙烯、反-2-丁烯、丁二烯、环氧乙烷以及氯乙烯等常见化学气体。图中结果表明，模型能够精确预测各气体浓度，充分体现出其较高的鲁棒性。

为了突显 Cross-talk Transformer 的显著优势，选取识别准确率、浓度预测 RMSE 及 MAE 三项指标，对比了 ACE、KNN、CNN、MLP 和常规 Transformer 模型，具体数值见表 2-4。

在实验数据集（含 11 种化学气体）的评估中，Cross-talk Transformer 的识别准确率达 98.63%，显著高于 ACE (93.60%)、KNN (97.12%)、CNN (98.08%)、MLP (98.25%) 及常规 Transformer (98.28%)。对于浓度预测任务，本模型的 RMSE 为 0.51 kppm，MAE 为 0.33 kppm，均明显低于其他方法，表明误差最小、精度最高。

在浓度预测任务中，对 Cross-talk Transformer、MLP 和 CNN 三种模型进行了深入对比，图 2-16 展示了它们在 11 种气体上的 RMSE 和 MAE 指标表现。RMSE 反映预测值与实际值之间的均方根误差，而 MAE 反映绝对误差的平均值，是衡量预测精度的重要依据。

从图中数据可见，对于甲烷、乙烯、氨气、六氟化硫、三甲胺、丙烯和氯乙烯等气体，Cross-talk Transformer 的 RMSE 明显低于 CNN，表明其在预测这些气体浓度时误差更小、精度更高。同时，与 MLP 相比，对于甲烷、六氟化硫、三甲胺、环丙烷和氯乙烯等气体，Cross-talk Transformer 的 MAE 也显著较低。

综合分析表明，Cross-talk Transformer 在大多数气体的浓度预测中都具备出

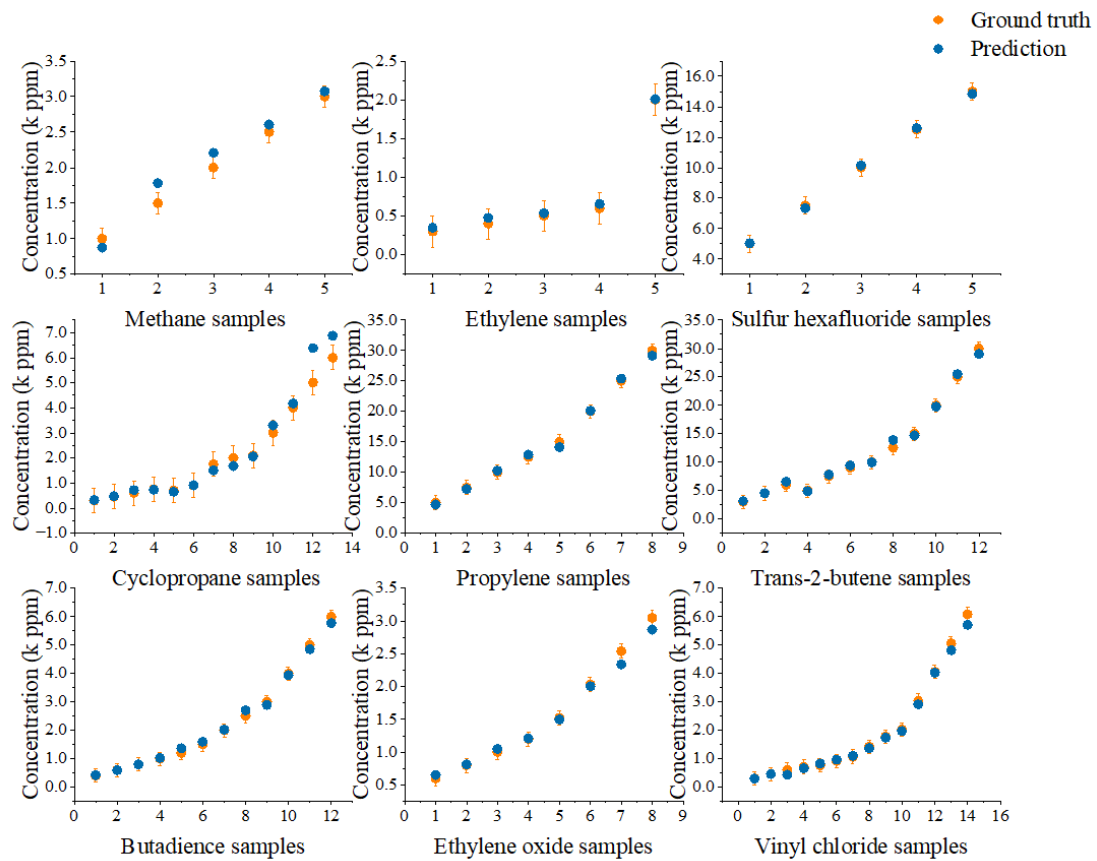


图 2-15 Cross-talk Transformer 在实验数据集上的浓度预测结果。
Figure 2-15 Concentration Prediction Results of Cross-Talk Transformer on the Experimental Dataset.

表 2-4 气体识别和浓度预测方法准确率比较。
Table 2-4 Comparison of Accuracy for Gas Identification and Concentration Prediction Methods.

气体识别方法						
Metric	ACE	KNN	CNN	MLP	Transformer	Ours
Accuracy	93.60%	97.12%	98.08%	98.25%	98.28%	98.63%
浓度预测方法						
Metric	CF	PMI	CNN	MLP	Transformer	Ours
RMSE(kppm)	1.26	0.77	0.53	0.58	0.54	0.51
MAE(kppm)	0.92	0.47	0.34	0.39	0.34	0.33

色的准确性，这进一步证明了该模型在工业环境化学气体监测中的优越性能和实用性。

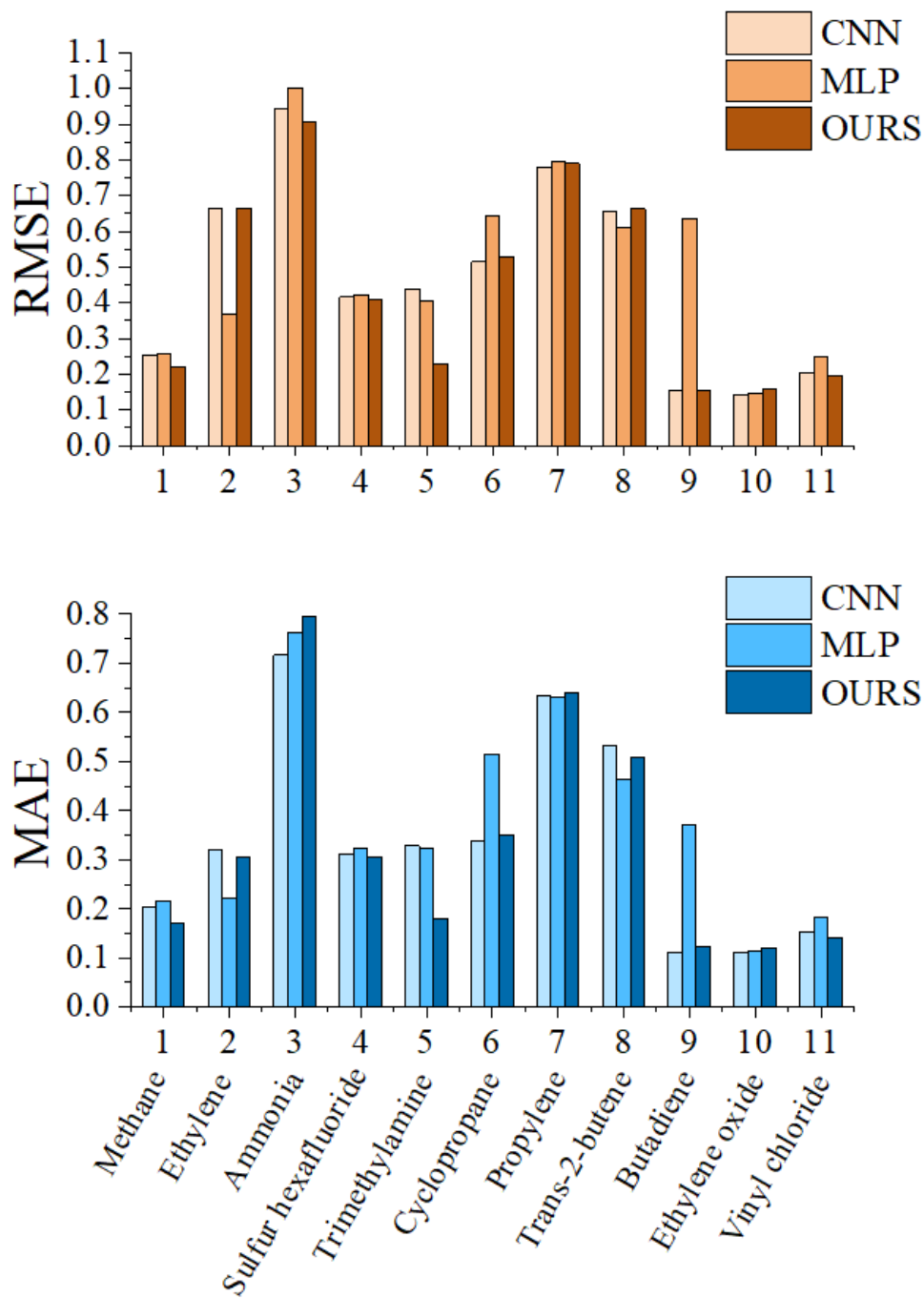


图 2-16 Cross-talk Transformer、CNN 和 MLP 在浓度预测任务上的比较。

Figure 2-16 Sample Figure

2.7 本章小结

本章主要讨论了长波红外（LWIR）光谱成像在各种应用中的关键作用，尤其是在气体监测领域。尽管先进的计算光谱成像技术已取得显著进展，但无冷却

快照红外光谱仪 (USIRS) 系统仍面临光谱分辨率低和数据噪声大等挑战。深度学习被看作是一种有前途的解决方案, 但其在 LWIR 成像中的应用仍受到长波高光谱图像数据严重缺乏的限制。

为解决这些问题, 本文提出了两种策略。首先, 提出了基于像素的 hierarchical spectral transformer (HST), 这是一种从公开的单像素长波红外光谱数据库中学习的深度学习框架, 旨在实现高光谱分辨率的 LWIR 光谱图像重建。实验结果证明了 HST 在提高光谱分辨率和减弱噪声方面的鲁棒性和有效性。其次, 提出了 cross-talk transformer, 这是一种新的算法, 旨在直接在低质量的数据上运行高鲁棒性的后端任务, 特别是在气体识别领域。实验结果表明, 这种方法在气体类型和浓度的监测方面具有高精度和强大稳健性, 证明了其在工业监测和其他应用中的潜力。

第3章 基于神经光谱场的混合监督学习光谱重建

3.1 引言

计算光谱成像对许多领域的发展至关重要,然而,在其物理实现中,常面临光谱分辨率低和编码设计受限的问题,这对光谱成像的能力造成了限制。而神经网络作为一种强有力的工具,通过引入光谱先验知识能有效提高光谱分辨率,已在高光谱成像领域取得了显著的成就。尽管如此,特定光谱成像仪器所能提供的数据不足,仍然束缚了神经网络在光谱成像中的应用,阻碍了鲁棒模型的建立。

针对此,本文提出了一种基于神经光谱场的混合监督学习策略来进行光谱重建。首先利用物理模型模拟得到的数据进行有监督学习。然后,再通过神经光谱场的建模方式优化神经网络输出的光谱曲线。最后引入一种基于物理的细化方法,以增强重构光谱与测量数据的一致性,从而实现高质量的光谱重建。

为验证提出的方法的有效性,选择了一个设计受限的微型计算光谱仪作为实验对象。此类光谱仪在生物医学分析、环境监测和食品安全检测等领域均有巨大的应用潜力,因此引起了学术界和产业界的广泛关注。其中,由 AlGaAs/GaAs 材料构成的带隙 PN 结光谱仪是一种基于标准 III-V 族半导体工艺制备的光谱探测器件。其便于工业化的设计和独特的电压可调谐光响应特性,都将推动微型光谱仪的商业化进程。为了利用这种独特的电压可调谐特性进行光谱检测,应用了提出的基于神经光谱场的混合监督学习逆向求解策略,并建立了波长-偏压双维响应矩阵以实现高效的光谱解析。测试结果表明,该光谱仪的波长精度可以达到 0.30nm ,光谱分辨率突破 10nm ,在成像性能测试中也显示出优秀的光谱成像能力。

3.2 计算微型光谱仪

3.2.1 计算微型光谱仪的设计

光学光谱仪一直是至今为止材料表征和化学分析的首要工具之一。传统的台式光谱仪通常依赖于体积庞大的色散光学元件、长光程、探测器阵列和可移动部件,这些需求阻碍了其在需要最小化体积、成本和功耗的应用中的小型化进程。通常,缩小光谱仪尺寸需要在其分辨率、动态范围或信噪比方面作出妥协。然而,可以设计微型光谱仪以达到特定应用的“可接受”性能水平^[84,85]。在许多情况下,如果目标是识别特征光谱峰而非精确测量,那么在可见光范围内的“满意”分辨率可能在 10nm 左右,甚至更大^[86]。此外,通过提升某个特定性能方面,微型光谱仪可以专门用于实现传统系统难以实现的极端测量。例如,片上单光子光谱仪^[87]和单纳米线光谱仪^[88]的优势分别在于其超高灵敏度检测和超紧凑的空间占用。

自 1990 年代初以来,已经有基于各种设计和工作原理的小型光谱仪,它们

具有一系列的工作光谱带和分辨率^[89]。然而，在过去十年中，“重构”或“计算”光谱仪系统利用了更显著的计算机处理能力以及微处理器尺寸和成本的降低^[90-94]。这些系统通常具有一组具有独特光谱响应特性的探测器，当并行测量时，可以使用复杂的算法结合起来，近似或“重构”入射光谱。这些系统不仅可以利用硬件技术的进步，还可以利用新计算方法的发展，特别是基于压缩感知和机器学习的方法。

在光谱仪中已经有一系列可调的窄带滤光片，如声光可调滤光片（Acousto-optic Tunable Filter, AOTF）^[95]、液晶可调滤光片（Liquid Crystal Tunable Filter, LCTF）^[96]、法布里-珀罗滤光片^[97-99]以及微环谐振器^[100]。它们的光谱传输可以通过施加电压或声信号进行快速和动态的控制，从而在时间上分离光谱成分。然而，至今为止，双折射晶体的尺寸限制已经成为 AOTF 光谱仪小型化的主要障碍。LCTF 遭遇了类似的限制，包含一系列偏振器和液晶单元，这些都难以小型化。另一方面，可调法布里-珀罗滤光片可以通过成熟的微机电系统（Micro-Electro-Mechanical Systems, MEMS）兼容工艺制造，使其非常适合微型光谱仪的大规模生产^[101,102]。

基于可调法布里-珀罗滤光片的微型光谱仪的典型结构具有一个由两个平行镜子组成的共振光学腔，镜子之间的距离可变。在给定设备的情况下，假设吸收率和反射率是固定的，传输光谱可以通过改变光程来调节，这可以通过改变光路长度、介质厚度或入射角来实现。调节镜子之间的距离是最常见的策略，通过使用静电或压电致动器实现，当前的 MEMS 技术可以轻易地提供广泛的调谐范围。折射率的调谐可以通过选择电光可调 LiNbO_3 ^[103] 或液晶^[104,105] 作为腔体介质来实现；入射角的调节可以通过旋转滤光片来实现^[106]。

在设计这些高性能的法布里-珀罗光谱仪时，需要考虑某些关键因素。传输峰的全宽半高（Full Width at Half Maxima, FWHM）决定了基于滤光片的光谱仪的光谱分辨率，而光谱分辨率等于腔体的固有精细度。因此，高分辨率需要高反射率，但对于金属镜腔体，这会导致更低的传输率，因此信噪比较低。部分解决方案是使用分布式布拉格反射镜作为镜子，这些反射镜由交替的高低折射率介电四分之一波层组成，具有高反射率和低吸收率在特定光谱范围，但它们成本高且制造复杂^[107]。需要注意的是，腔体系统中的缺陷，包括不平行或镜面缺陷，也会降低有效精细度，因此降低分辨率。

2024 年，Wang 等人提出了一种基于 $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ 材料的渐变带隙 PN 结光电探测器，其电压可调响应可作为光谱仪使用^[30]。与侧向工作的梯度材料纳米线不同，渐变带隙 PN 结光谱仪依赖于纵向堆叠的组分梯度外延结构，并通过改变 PN 结的耗尽厚度实现电压可调响应。渐变带隙 PN 结光谱仪的设计依赖于纵向外延结构梯度，这意味着设备的物理横向尺寸不影响其功能。这种光谱仪可以通过标准 III-V 工艺制造，其超小的占地面积可缩小到微米级，并且具有广泛的光谱范围，从 480nm 到 820nm。但渐变带隙 PN 结光谱仪的光谱响应依然受到与其结构高度相关，因此依然受到物理实现的约束。因此如何从光谱解码角度

出发实现高准确度的光谱重建是至关重要的问题。

3.2.2 微型光谱仪的工作机制

传统 PN 结光电二极管是一种典型半导体器件，其工作原理基于入射光子在器件内产生的光生电子—空穴对，这些载流子在耗尽区电场作用下通过扩散或漂移被收集。其中，耗尽区由 P 区与 N 区因载流子扩散而形成的空间电荷区构成；扩散源于浓度梯度，而漂移则受电场驱动。

随着加在 PN 结两端（N 区正极、P 区负极）的反向偏压增大，载流子在复合前的收集效率提高，从而改善了量子效率和响应度。量子效率衡量的是入射光子转化为电流的效率，而响应度则反映了光电二极管对特定波长光的敏感程度。然而，由于不同反向偏压下的响应—波长曲线呈高度线性关系，单纯调节偏压无法获得更丰富的光谱信息。因此，传统 PN 结光电二极管并不适用于直接构建光谱仪，后者要求对不同波长光具有差异化响应。

如图3-1所示，在渐变带隙 PN 结的设计中，通过在 N 型区域采用成分梯度的 $Al_xGa_{1-x}As$ （其中 x 由 0.5 逐渐降低至 0）实现带隙调控。结果是，接近 PN 结处具有较宽带隙的 $AlGaAs$ ，而远离结区则呈现较窄带隙的 $AlGaAs$ 或 $GaAs$ 。在低反向偏压下，仅宽带隙层中的载流子（主要由短波长光激发）能被有效收集，因为窄带隙层因价带势垒限制长波长光激发的空穴贡献。随着反向偏压增大，耗尽区延伸至窄带隙区域，从而使得长波长入射光产生的载流子也能参与收集并贡献光电流。此调控机制实现了电压可调、波长选择性响应，为光谱仪的设计提供了新思路。

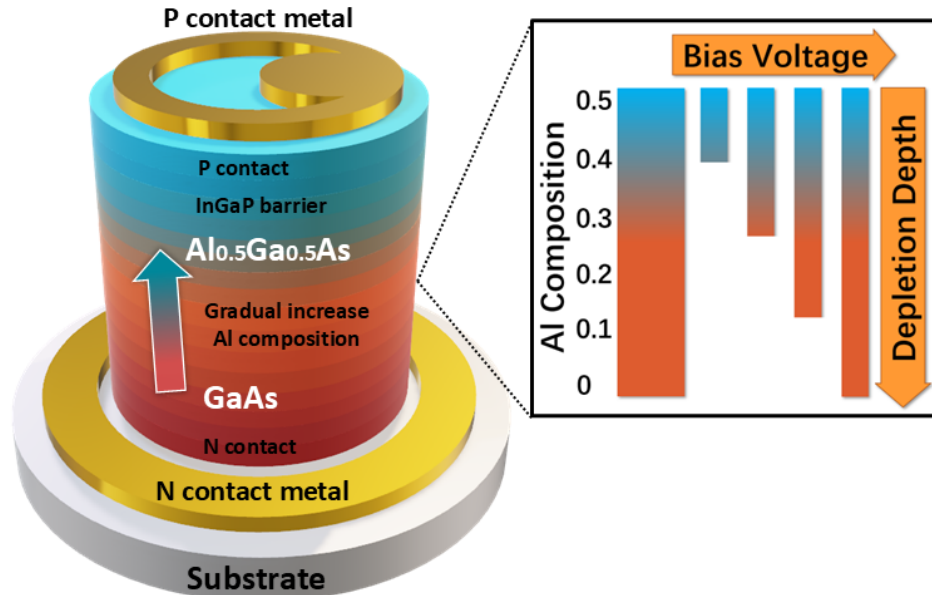


图 3-1 渐变带隙 PN 结光谱仪的示意图和电性能。

Figure 3-1 Schematic and electric properties of the p-graded-n junction spectrometer.

电压可调的光谱响应对后续的光谱重建至关重要。光谱重建指的是将光谱

数据转化为图像或其他可视化形式的过程,有助于深入理解和利用光谱信息。因此,渐变带隙 PN 结的设计为光谱重建提供了新的机遇。

该渐变带隙 PN 结光谱仪具有以下显著优势:首先,它能够通过单像素实现高分辨率的波长测量,大幅提升了测量效率,相比传统依赖多个探测器或精密光栅的设计,系统更为简洁;其次,该结构不局限于特定材料,可采用不同带隙的半导体系统,因而在近红外或中波红外等波段的应用中也具备潜力;再次,该器件可借助倒装芯片键合技术集成于焦平面阵列,实现背照明,从而同时获得图像和光谱信息,拓展了在遥感、医疗诊断和工业检测等领域的应用范围。

在器件结构方面,所有外延层均与 $GaAs$ 衬底晶格匹配,以确保材料的结构质量和性能。生长工艺始于一层厚度为 500 nm 、掺杂浓度为 $2 \times 10^{18}\text{ cm}^{-3}$ 的 $n+$ $GaAs$ 接触层,随后沉积一层厚度 $2\text{ }\mu\text{m}$ 的渐变 n 型 $Al_xGa_{1-x}As$ 层,其中 Al 组分由 0 逐渐增加至 0.5 ,实现带隙梯度调控。在 n 型 $Al_{0.5}Ga_{0.5}As$ 层中插入一层 50 nm 的 $InGaP$ 空穴阻挡层,以抑制由长波长光激发的空穴侵入耗尽区。接下来是 p 型 $Al_{0.5}Ga_{0.5}As$ 层和顶部的 $GaAs\ p+$ 接触层,最后在最外层沉积 $Ti/Pt/Au$ 金属层,确保电信号的高效传输。

该结构通过在 PN 结附近设置较大带隙材料,实现了电压可调的响应特性。短波长光在高 Al 组分区域吸收,而长波长光则较易穿透至低 Al 组分区域。在低反向偏压下,由于价带势垒限制,长波长光产生的空穴难以扩散至耗尽区,因此对光电流贡献较小;而提高反向偏压使低 Al 组分区域耗尽,从而使少数载流子得以收集。这样,入射光谱信息得以编码在光电流—电压特性中,实现了高效的光谱测量。

此外,将较大带隙材料置于结区有助于在高反向偏压下显著降低隧穿电流,这对短波红外 (Short Wavelength Infrared, SWIR) 和中波红外 (Mid Wavelength Infrared, MWIR) 探测器尤为重要。通过这一设计,可更好地满足长波段光电探测器对性能和稳定性的要求。

3.2.3 微型光谱仪的电学性能

器件的暗电流和光电流密度曲线展示了其出色的性能特性。在室温环境下,器件在 -10 V 的反向偏压下表现出极低的暗电流密度,其值只有 $3.47 \times 10^{-9}\text{ A/cm}^2$,这意味着在没有光照射时,器件的电流产生非常小,从而可以保证其高的信号噪声比。此外,器件在 480 nm 至 820 nm 的工作波长范围内均有响应,证明了其优秀的光谱测量能力。特别是在 775 nm 处,器件达到了峰值响应度,其值高达 0.51 A/W ,同时对应的外量子效率达到了 81% ,表明器件在捕获和转换光信号方面具有高效的能力。同时,探测器对应的热噪声限制探测率为 $1.86 \times 10^{13}\text{ cm} \cdot \text{Hz}^{1/2}/\text{W}$,表明在器件中,热噪声的影响被有效地控制在了一个低水平。得益于采用了成熟的 III-V 族半导体技术,光电探测器在性能上显著优于之前报道的单像素光电探测器光谱仪。

如图3-2所示,在光谱仪应用中,器件展现了电压可调的响应特性。随着反

向偏压增加，截止波长延长，原因在于新出现的耗尽区扩展至较低带隙 $AlGaAs$ 区域，使得长波长光得以响应。采用响应度随波长与反向偏压变化所构成的光谱响应矩阵进行光谱重建，不仅提升了器件性能，也为光谱重建提供了更多可能性和便利性。

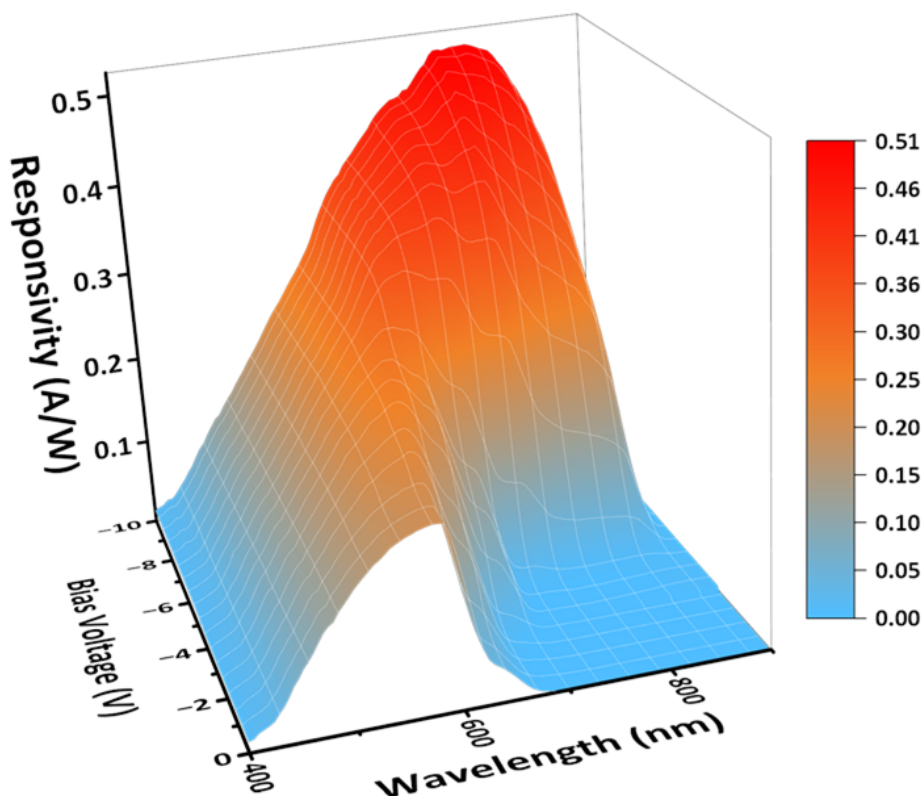


图 3-2 渐变带隙 PN 结光谱仪的示意图和电性能。

Figure 3-2 Schematic and electric properties of the p-graded-n junction spectrometer.

如图 3-3 所示，所制备的渐变带隙 PN 结光谱仪显微镜图像展示了器件的尺寸和形状细节。台面直径从 $20\mu m$ 到 $500\mu m$ 不等，使得该设计能够适应多种应用需求。器件尺寸主要受光刻机精度限制，因此通过调整光刻工艺可以灵活改变器件尺寸，以满足不同场景的需求。例如，对于要求高精度和大感光面积的应用，如手持或实验室光谱仪，可采用大尺寸设计；而对于集成到智能手机中等对尺寸要求严格的场合，则可制作微小光谱仪。值得注意的是，设计中最小台面直径为 $20\mu m$ ，而目前智能手机摄像阵列单个像素约为 $5\mu m$ ，表明该设计能够满足大多数移动设备的尺寸要求。

为最大化入射光与活性区域的耦合，选用了台面直径为 $180\mu m$ 的器件进行测量，此尺寸在耦合效率和操作便捷性之间达到了良好平衡。事实上，渐变带隙 PN 结光谱仪的功能与器件尺寸无直接关联，只要从台面顶部照射入射光，器件均可正常工作，这赋予了设计较高的灵活性。然而，将器件直径缩小至微米级以下可能导致光学耦合难度增加和对准公差问题，从而影响性能。因此，在设计中

需在器件尺寸和操作便利性之间权衡选择。

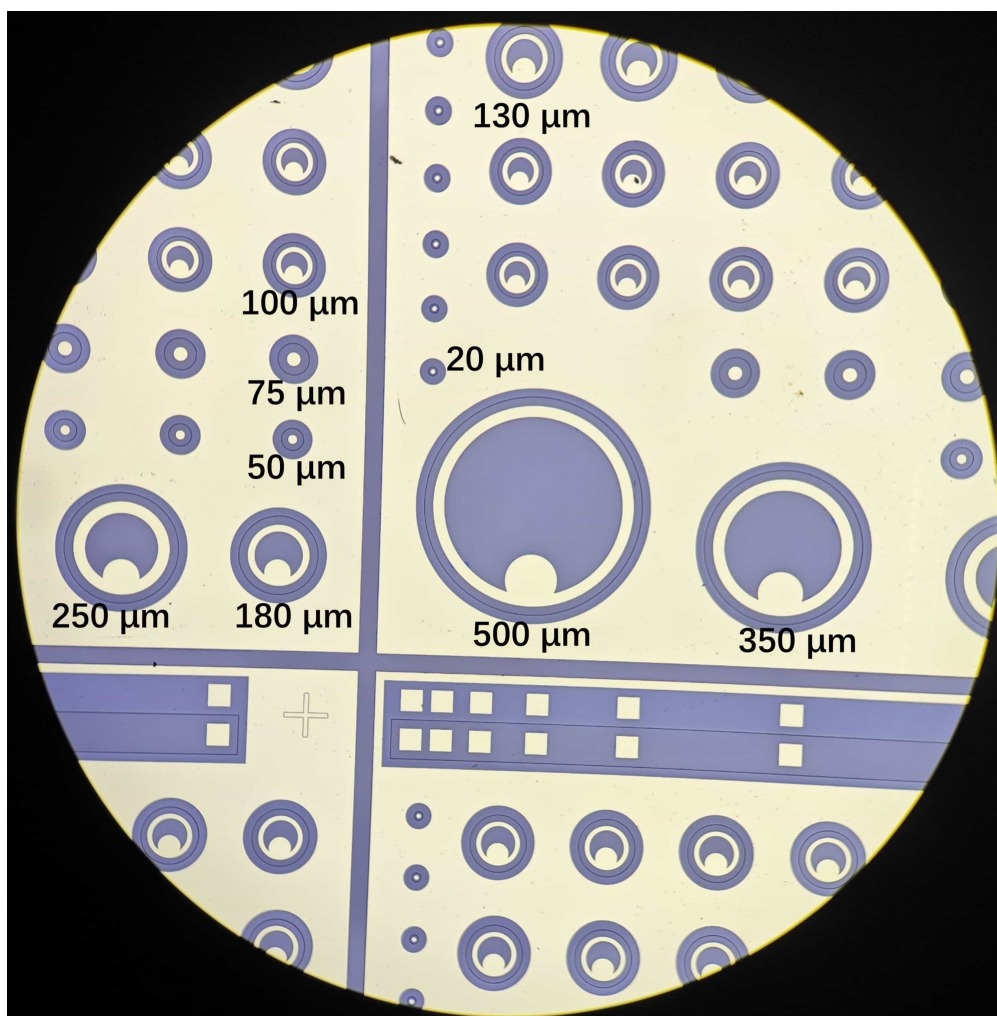


图 3-3 渐变带隙 PN 结光谱仪的显微镜图像

Figure 3-3 Microscope image of the p-graded-n devices with different footprint

3.3 光谱测量流程

如图 3-4所示，采用窄带单峰激光、窄带双峰激光和宽带光源（分别用暗红、蓝、绿色表示）对光谱仪进行测试。利用自由空间耦合方式，将光源输出的光谱信号导入光谱仪，入射光激发器件产生光电流。值得注意的是，施加不同偏压时，光谱仪的响应函数发生变化，从而导致输出光电流随之变化。通过调节偏压，实现入射光谱与光谱响应的结合，可得到反映器件性能的光电流曲线。

在给定偏压 V 下，光电流的测量可用下式积分方程描述，将未知入射光谱 $P(\lambda)$ 与预先校准的响应函数 $R(\lambda, V)$ 关联：

$$I(V) = \int_{\lambda_{\min}}^{\lambda_{\max}} R(\lambda, V) P(\lambda) d\lambda \quad (3-1)$$

其中, $I(V)$ 为测得的光电流曲线, 而 $R(\lambda, V)$ 描述了光谱仪在特定偏压下对各波长 λ 的响应; $P(\lambda)$ 表示未知入射光谱, 积分上下限 $\lambda_{\min}$ 和 $\lambda_{\max}$ 定义了仪器的工作波长范围, 超出该范围的波长难以获得准确测量。通过调节偏压获得一系列光电流曲线, 并利用相应数学工具解决这一逆问题, 最终重建出原始入射光谱。

如图 3-4所示, 展示了光谱仪在四个不同偏压下的光谱响应函数, 体现了其电压可调特性。从图中可观察到, 响应函数随波长增长而逐步展开, 通过调控偏压, 可以改变器件对各波长光的响应, 从而为从光电流曲线中重建原始入射光谱提供了可能。

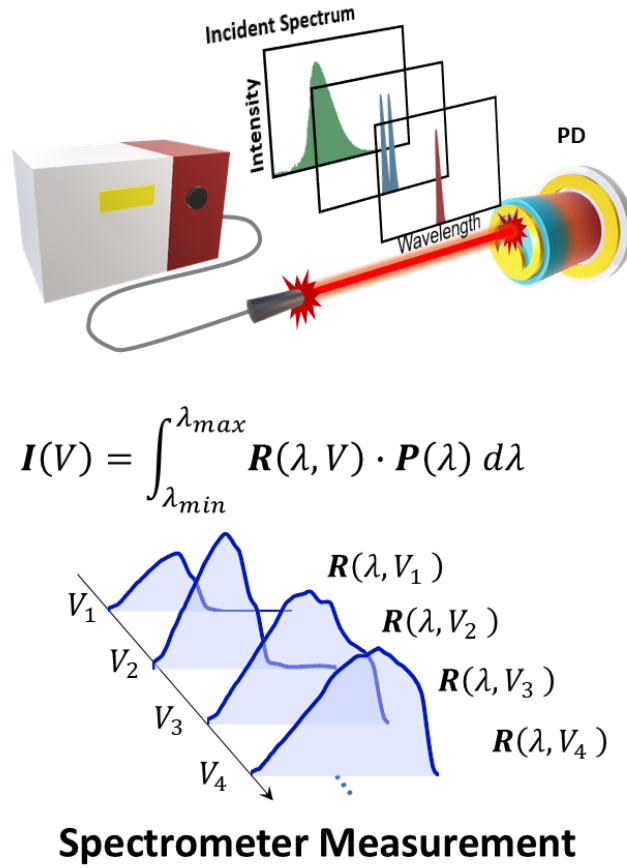


图 3-4 渐变带隙 PN 结光谱仪的光谱测量

Figure 3-4 Pipeline of the p-graded-n spectrometers

如图 3-5所示, 三种不同入射光源产生了各异的光电流曲线。曲线中的拐点反映了电压累积光谱响应的独特特征, 而这些特征正是光谱重建的重要依据。由于其高阶与非线性特性, 传统光谱重建算法难以直接处理该数据。为此, 引入神经网络方法, 通过训练使其自动捕捉复杂特征, 实现从光电流曲线到入射光谱的有效重建。

如图 3-6所示, 提出的神经网络方法采用三个主要模块: 特征提取器、神经光谱场和物理优化模块。首先, 通过物理模型仿真与光谱数据库构建合成数据

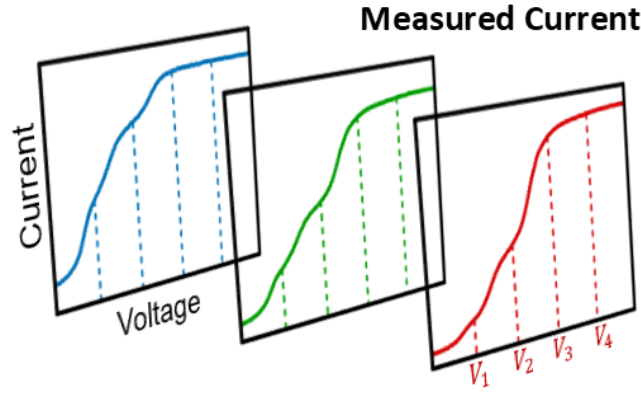


图 3-5 渐变带隙 PN 结光谱仪的光电流

Figure 3-5 Photon current

集，获得大量配对的光电流与入射光谱样本，为网络训练提供丰富信息，使其能够捕捉光电流与入射光谱之间的复杂映射关系。

特征提取器从输入光电流曲线中提取统计属性及电压累积光谱响应产生的高阶、非线性特征；神经光谱场由多层网络组成，将提取的特征映射到入射光谱空间，实现复杂映射学习；物理优化模块则利用对光电流物理模型的认识，对神经网络输出进行校正优化，从而进一步提高光谱重建的准确性。

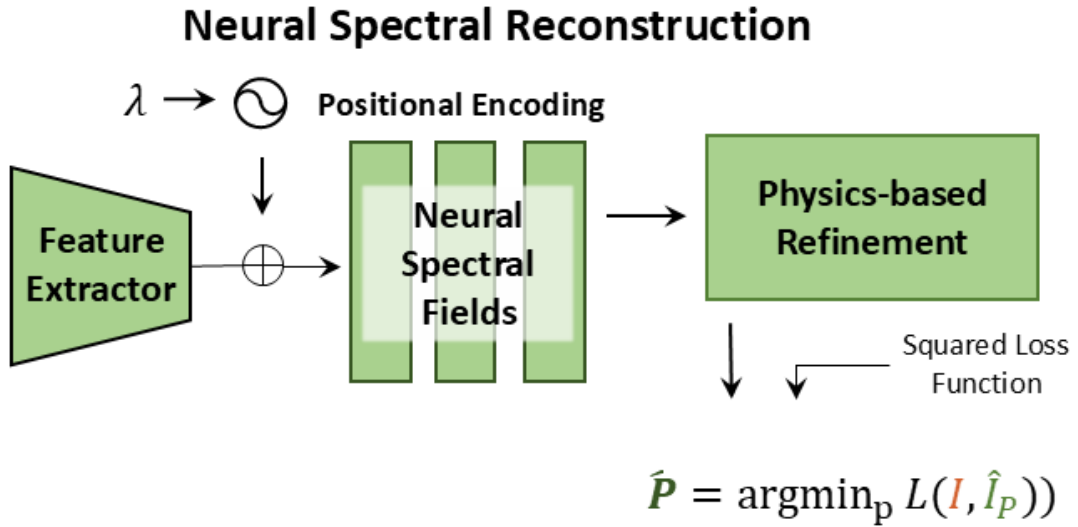


图 3-6 神经光谱场

Figure 3-6 Neural spectral fields

特征提取器的核心功能在于将输入的光电流数据经过一系列复杂的运算转换为有意义的特征表示，从而支持后续的数据识别和处理。为此，采用了多层深度学习框架，在每一层中，全连接网络按照权重对输入进行加权求和，并借由激活函数非线性转换，逐步提取出更抽象、更具表征意义的特征。这一逐层的抽象过程不仅帮助深入理解光电流数据的内在属性，还使得模型能够有效处理复杂非线性问题，并从大量数据中筛选出关键特征，从而整体提升网络性能与效能。

图 3-7展示了经过特征提取后不同类型光谱电流数据的可视化结果。具体而言，利用主成分分析技术将原始高维数据降至三维后，不同特征以颜色区分：红色对应单峰光谱数据，蓝色代表双峰光谱数据，绿色则指示宽带宽度特征。为了更加直观地观察光电流特征的分布，将这三类特征分别投影到三个独立平面上展示，能够观察到三类不同的光谱的特征在可视化特征空间中是可区分的。

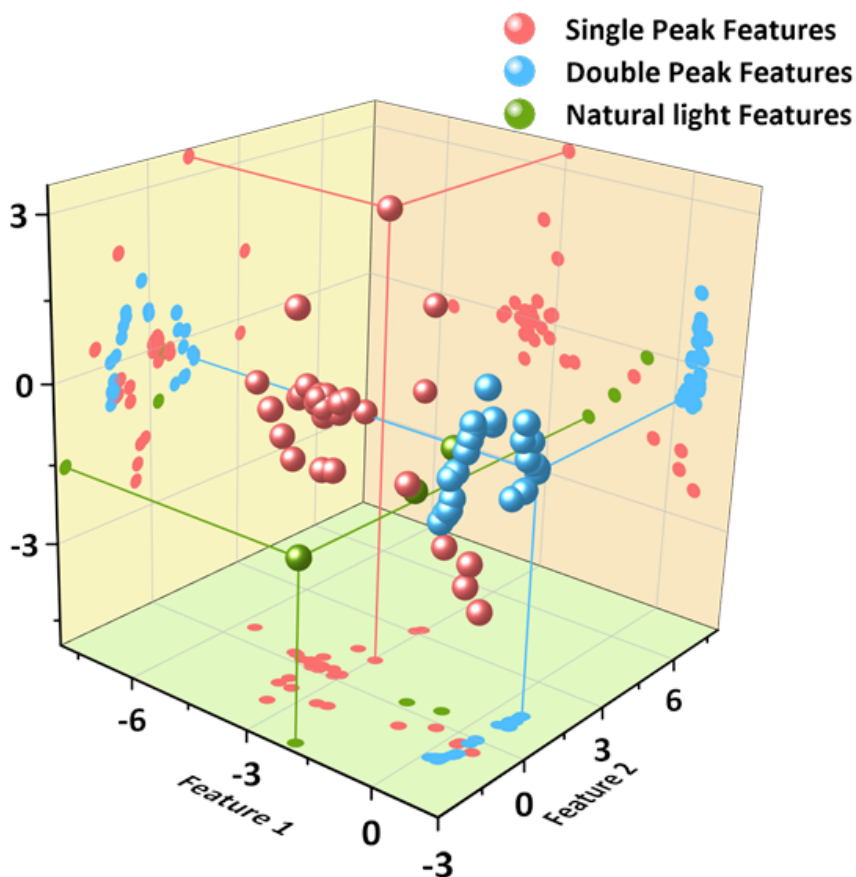


图 3-7 提取特征的低维分析

Figure 3-7 Low dimensional analysis

在进一步深入分析提取的特征时，引入了“神经光谱场”机制，将特征解码为直观的光谱曲线。首先，针对特定波长 λ 进行“位置编码”，通过一系列数学运算为每个 λ 赋予嵌入式高维表示，使得神经网络能够更准确地控制和理解光谱结构。随后，将该位置编码与特征提取器输出的丰富特征结合后，一同输入到线性层神经网络中。网络中每个节点通过加权处理，为每个波长 λ 计算出相应的光谱强度。

对关注区域内均匀线性分布的各个波长执行上述操作，确保获得平衡、充分采样的光谱数据，最终生成了一条包含大量关键点的初步光谱曲线。

神经光谱场解码出的光谱曲线在一定程度上能反映入射光特性，但可能未能完全满足由物理定律（如式3-2）导出的约束。为提高精度并确保与物理模型的一致性，对初步光谱曲线进行了进一步优化，即在曲线上依次实施平移、缩放

和线性强度变换。平移操作调整曲线在坐标轴上的位置；缩放操作改变曲线的整体形状和幅度；线性强度变换则调节曲线的局部强度，使其与测量数据更好匹配。

这些变换均通过系统化的组合优化实现。首先采用随机搜索策略，探索各类变换组合，再引入马尔科夫链蒙特卡罗（Markov Chain Monte Carlo, MCMC）方法，在概率框架内寻找最优解。经过这一系列精细调控后，得到的重建光谱曲线在内容和形式上与实际光电流数据具有高度一致性。

图 3-8 直观展示了三种不同类型的入射光谱重建结果，为深入解析光谱特性提供了支持。单峰光谱的结构简单，中心波长定于 600nm。单峰光谱在中心波长处呈现突出的峰值，并随波长两侧迅速衰减，体现出清晰的能量集中效应。双峰光谱表现出两个峰值，中心波长分别位于 580nm 和 590nm。中心波长接近的两个峰值反映了光谱重建的困难。宽带光谱同样以 600nm 为中心，但与单峰光谱相比，其能量分布更加平缓宽广，表明在较大波段内具有均匀的响应，对重建方法提出更高的要求。

这些结构各异的光谱重建结果，将用于验证了所提出的光谱仪和神经网络重建方法的有效性和实用性，而且展示了该技术在应对不同场景和需求时的鲁棒性与普适性。

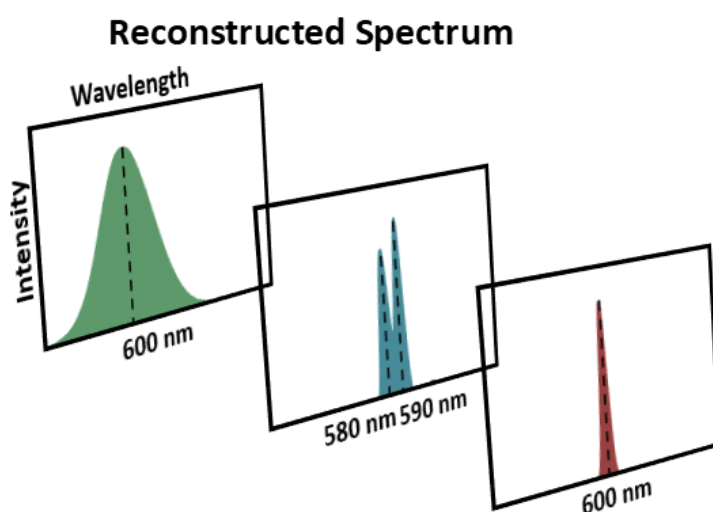


图 3-8 重建光谱

Figure 3-8 Reconstructed spectrum

3.4 光谱重建

3.4.1 重建问题描述

在利用提出的光谱仪测量未知光谱时，对于任一特定偏压 V ，测量到的光电流 I 可由下式给出：

$$I(V) = \int_{\lambda_{\min}}^{\lambda_{\max}} R(\lambda, V) P(\lambda) d\lambda \quad (3-2)$$

其中, $I(V)$ 表示在偏压 V 下得到的光电流曲线, 该曲线是通过在一系列不同偏压下测量获得的, 反映了光谱仪的重要可调谐特性。积分区间 $(\lambda_{\min}, \lambda_{\max})$ 定义了光谱仪的工作波长范围; $R(\lambda, V)$ 是预先校准得到的光谱响应函数, 它表征了仪器在不同波长和偏压下的响应特性; 而 $P(\lambda)$ 则是通过实验需要重建的未知入射光谱。

在实际部署中, 会向设备施加一系列离散偏压 $\{V_1, V_2, \dots, V_M\}$, 并在每个 V_i 下测得对应的光电流 I_i ($i = 1, 2, \dots, M$), 从而获得光电流序列 $\{I_1, I_2, \dots, I_M\}$ 。同时, 将未知光谱 $P(\lambda)$ 离散化为长度为 N 的序列 $\{P_{\lambda_1}, P_{\lambda_2}, \dots, P_{\lambda_N}\}$, 其中每个 P_{λ_i} 对应波长 λ_i 下的光谱值。

基于上述离散化过程, 可得以下离散形式:

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_M \end{bmatrix} = \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,N} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M,1} & R_{M,2} & \cdots & R_{M,N} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{bmatrix} \quad (3-3)$$

可以简化为:

$$I = RP \quad (3-4)$$

尝试采用了多种经典优化数值方法作为对比。首先, 采用均方误差来最小化预测值与真实值之间的平方误差均值, 从而指导模型参数更新, 提升整体性能。但在处理某些特殊光谱重建任务时, MSE 显得不足以满足需求。接着, 引入了 Tikhonov 正则化, 通过在损失函数中加入正则项, 将模型复杂性纳入考量, 有效缓解了过拟合问题并增强了解的稳定性。然而, 对于本设备特有的“电压累积”特性, Tikhonov 正则化难以充分利用光电流的特征。考虑到数据稀疏性的重要性, 进一步采用了稀疏 L1 正则化, 该方法在高维数据的特征选择中表现优异, 但在实际应用中, 也未能充分挖掘设备的独特优势。综合审视这些传统方法的优缺点后, 鉴于实验已验证设备特有的“电压累积”特性, 创新性地提出了一种全新的解决方案——利用神经光谱场算法实现光谱重建。该方法针对设备的内在特性进行量身定制, 在数据处理中充分发挥设备潜能, 实现了更高效、更精确的光谱重建。

3.4.2 光谱重建流程

在通过光谱仪测量获得光电流曲线后, 下一步是光谱重建过程, 准确获取未知入射光谱。整个重建流程大致可划分为以下五个步骤:

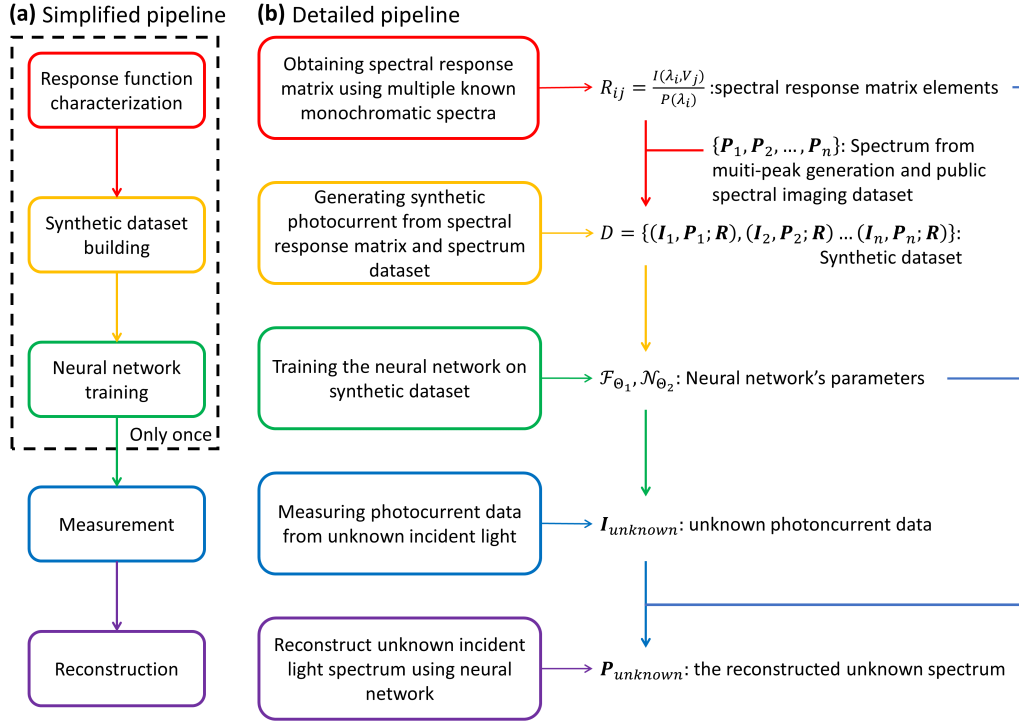


图 3-9 渐变带隙 PN 结光谱仪的光谱重建流程

Figure 3-9 Pipeline diagram of the p-graded-n spectrometers

1. 定量化光谱响应函数：首先，需要对光谱响应函数进行定量化。该响应函数以矩阵形式表达，其中每个元素反映了特定单色光入射下所产生的光电流。为此，利用单色仪产生多个已知单色光（记为 $P(\lambda_i)$ ），在每个设定偏压 V_j 下测量对应的光电流 $I(\lambda_i, V_j)$ ；从而可以通过 $R_{ij} = \frac{I(\lambda_i, V_j)}{P_i}$ 得到响应函数各项。该步骤能够准确刻画设备特性，为后续未知光谱的重建提供物理基础。

2. 构建神经网络训练数据集：利用定量化的光谱响应函数和庞大的已知光谱库，构建了用于神经网络训练的方阵数据集。设光谱库表示为 $\{P_1, P_2, \dots, P_n\}$ ，参考器件响应模型，生成模拟数据集 $D = \{(I_1, P_1; R), (I_2, P_2; R), \dots, (I_n, P_n; R)\}$ 。该大规模数据集为后续神经网络学习光电流与入射光谱之间的复杂映射关系提供了充足的支持。

3. 神经网络训练：接下来，通过训练神经网络实现从光电流到光谱数据的映射。在网络中，将关键参数分别记为 F_{Θ_1} （前端卷积部分）和 F_{Θ_2} （条件神经光谱场部分）。整个训练过程可以看作求解一个复杂的最优化问题，需要依赖大规模数据集和高效的优化器不断调整网络参数，以准确捕捉光电流与入射光谱之间的关系。

4. 实际光谱测量：随着神经网络训练完成，便可进入实际应用阶段。使用光谱仪测量未知光谱 $P_{unknown}$ 时，在各个预设偏压下精确记录光电流 $I_{unknown}$ ，形成完整光电流曲线。该曲线包含了所有重建未知入射光谱所需的关键信息。

5. 光谱重建：最后，将测得的光电流曲线作为输入，利用已训练的神经网络对未知光谱进行重建。神经网络利用此前学习到的映射关系，对输入数据进行解析和推断，输出接近真实 $\mathbf{P}_{\text{unknown}}$ 的重建光谱。此过程实现了高精度、高效率的光谱重建，且在完成初期准备工作（网络设计、数据构建和训练）后，实际测量与重建步骤简单而便捷，可按需求反复执行。

3.4.3 响应函数表征

为了保证后续步骤中获得准确的光谱重建结果，在实验前对光谱响应函数 R_{ij} 进行校准。具体流程如下：

首先，使用单色仪生成带宽约 2.31 nm 的准单色光。利用透镜和光纤将光照射到设备上，激发电流。为确保实验在标准化环境中进行，选定入射光波长范围为 400 nm 至 900 nm，并对单色仪输出的光功率相对于标准响应度探测器进行归一化处理，从而消除外部干扰影响。同时，为减少电路干扰，提高测量精度，选择设备偏压范围为 0 V 至 -10 V。

在光谱响应矩阵的测量过程中，采用了以下设置：首先，用 180 Hz 的斩波器对入射光信号进行调制，使连续光源转变为脉冲信号，从而更好地控制光强；其次，将调制信号送入锁定放大器，该设备能在强背景噪声中提取出微弱的信号；再次，利用源表对光电探测器进行精确偏置，确保设备稳定运行；最后，采用卤素灯作为宽带光源，通过单色仪（型号 iHR 320，由 Horiba Instruments 生产）生成准单色光，以保证测量过程中光斑尺寸为约 50 μm ，且单色光的半高全宽（Full-width at the half of the maximum, FWHM）大约为 1 nm，从而达到高精度测量要求。

在此基础上，每个光谱响应矩阵元素 R_{ij} 定义为 $R_{ij} = \frac{I(\lambda_i, V_j)}{P(\lambda_i)}$ 其中 $I(\lambda_i, V_j)$ 表示设备在特定波长 λ_i 和偏压 V_j 下产生的光电流，而 $P(\lambda_i)$ 表示该波长下的入射光强度。校准后的响应函数以矩阵形式存储，作为后续光谱重建的重要物理参数。

此外，为进一步确保设备响应度的可靠性，还采用了具有 NIST 可溯源校准的 Thorlabs FDS-1010 Si 探测器进行最终校准。通过以上一系列校准步骤，能准确获得设备的响应矩阵，为后续基于神经网络的光谱重建提供了依据和数据支持。

3.4.4 神经网络结构

3.4.4.1 特征提取器

在本研究中，采用多层感知器作为特征提取器，其主要任务是从测量的光电流中提取特征。本研究构建的 MLP 包含 12 个隐藏层，每层配备 1024 个神经元，并采用 Leaky ReLU 激活函数以增强网络的非线性表达能力。同时，在每个隐藏层后均附加了批量规范化层和跳跃连接，以加速训练过程并有效防止过拟合。

在特征提取过程中，设定的特征维度为 501。该维度经过实验验证，既能充

分捕获电流信号中的主要信息，又不造成过多计算负担。提取出的特征经过可视化处理后，如图 3-7 所示，显示出单峰特征与双峰特征在特征空间中具有明确区分，而宽带自然光特征的分布则明显不同。这种分布特性为神经光谱场在光谱重建中实现对不同光谱的精确区分提供了有力支撑，也是本研究选择多层感知器作为特征提取器的主要依据。

3.4.4.2 神经光谱场

在本文的研究框架中，神经光谱场（Neural Spectral Fields, NSF）用于根据提取的特征和给定波长重建光谱。首先，对输入波长进行位置编码处理，以增强网络对高频光谱的表示能力。具体而言，位置编码由下式定义：

$$\gamma_{pos}(\lambda) = [\sin(2^0 \pi \lambda), \cos(2^0 \pi \lambda), \dots, \sin(2^n \pi \lambda), \cos(2^n \pi \lambda)] \quad (3-5)$$

其中， λ 为原始波长， n 为位置编码维度，在本实验中取 $n = 5$ 。通过正弦和余弦函数对波长进行一系列尺度变换，将原始波长嵌入到新的高维特征空间中。

经位置编码处理的波长与已提取特征一起输入到一个小型 MLP 中，该 MLP 包含 3 个隐藏层，每层设有 256 个神经元。经过该网络处理后，输出即为在输入波长处的光谱强度，构成连续光谱的查询结果。NSF 的核心功能在于根据位置编码和特征信息，为给定波长生成对应的光谱强度。

3.4.4.3 基于物理的细化

尽管 NSF 输出的光谱大体上能反映实际光谱，但其结果并不能与实际测得的电流完全对应。为进一步提升重建光谱的准确性，引入了基于物理模型的细化算法，对 NSF 输出进行迭代调整。在该细化过程中，每次迭代中随机将输出光谱平移约 1 nm，并对振幅进行缩放（因子分别取 1.1 与 0.9），以便在一定范围内产生微调，进而使对应光电流与实际测量值之间的误差最小化。针对每个输入电流均重复此细化过程 200 次，并最终选取误差最小的光谱作为最终重建结果。尽管该过程增加了计算复杂度，但其显著提高了光谱重建的精确性。

3.4.5 合成数据集构建

本研究构建了一个合成数据集，用于充分训练神经网络，该数据集包含三部分：光谱 $\mathbf{P}$ 、电流 $\mathbf{I}$ 和响应函数 $\mathbf{R}$ 。其中， $\mathbf{P}$ 表示光谱； $\mathbf{I}$ 是光电流； $\mathbf{R}$ 是系统的光谱响应矩阵。

合成数据集的光谱来源分为两类。第一类采用高斯多峰函数对激光光谱进行模拟，其表达式为

$$\mathbf{P} = \sum_{i=1}^2 A_i \exp\left\{-\frac{(\lambda - \lambda_i)^2}{2\sigma_i^2}\right\} \quad (3-6)$$

其中, A_i 表示各峰的高度, λ_i 为峰的中心波长, σ_i 则反映峰的宽度。在单峰数据生成中, 设定 $A_1 = 1.0$ 、 $A_2 = 0.0$, 同时令 λ_i 均匀分布于 $[480 \text{ nm}, 820 \text{ nm}]$, σ_i 均匀分布于 $[5 \text{ nm}, 100 \text{ nm}]$; 而在双峰数据生成中, A_1 与 A_2 均匀分布于 $[0.8, 1.2]$, λ_i 同样位于 $[480 \text{ nm}, 820 \text{ nm}]$, 但 σ_i 限定在 $[5 \text{ nm}, 20 \text{ nm}]$ 区间。更复杂的多峰情况 (如三峰函数重建, 参见第 3.5.5 节) 也可纳入模拟数据集中, 以确保复杂光谱的准确重建。

第二类光谱数据来源于自然光, 选取了广泛认可的 ICVL (Imperial Computer Vision & Learning) 数据集中包含校准板的实验室场景数据^[108]。在处理光谱图像时, 首先依据校准白板的标准进行像素级校准, 以保证所提取光谱信息的精度。

在获得光谱数据后, 利用经过精心校准的响应函数 $\mathbf{R}$ (详见第 3.4.3 节) 对离散光谱仪模型 (参见方程 3-4) 进行操作, 从而计算出对应的电流 $\mathbf{I}$ 。为模拟实际测量中的误差, 在 $\mathbf{I}$ 中引入了 0.1% 的高斯噪声, 尽管这种噪声不可能完全复制真实误差, 但可有效提升模拟结果的可信度。

最后, 利用数据集 $\mathbf{D} = \{(\mathbf{I}_1, \mathbf{P}_1; \mathbf{R}), (\mathbf{I}_2, \mathbf{P}_2; \mathbf{R}), \dots, (\mathbf{I}_n, \mathbf{P}_n; \mathbf{R})\}$ 对神经网络进行训练, 实现额外的光谱重建。

3.4.6 神经网络训练

本文提出的重建算法包括两个神经网络模型。第一个模型为特征提取器, 记作 $\mathcal{F}_{\Theta_1}$, 其中 $\mathcal{F}$ 表示特征提取功能, Θ_1 为相应网络参数, 能够从输入电流-电压曲线中挖掘深层特征。第二个模型为神经光谱场, 记作 $\mathcal{N}_{\Theta_2}$, 其中 $\mathcal{N}$ 表示生成光谱场的功能, Θ_2 则代表该网络参数, 通过优化使其根据提取的特征输出对应的光谱场。

算法将输入数据处理分为两步: 首先由特征提取器提取特征, 其次由神经光谱场根据这些特征生成光谱。这一分解策略既提高了算法灵活性, 又保持了模型结构的简洁性, 有助于直观解释实验结果。

初步阶段, 在构建的数据集上训练特征提取器, 目标是使其能够从无明确标签的电流-电压曲线中提取深层特征。为解决数据集中缺乏显式特征的问题, 采用联合训练方法, 同时优化 $\mathcal{F}_{\Theta_1}$ 和 $\mathcal{N}_{\Theta_2}$, 利用真实光谱作为监督信号, 通过比较联合输出与真实光谱的差异来更新参数, 其优化目标为

$$\min_{\Theta_1, \Theta_2} \sum_{i=1}^n \|\mathcal{N}_{\Theta_2}(\mathcal{F}_{\Theta_1}(\mathbf{I}_i)) - \mathbf{P}_i\|_2^2 \quad (3-7)$$

训练过程中, 执行了 100,000 个 epoch, 每个批次含 1,000 个样本。为确保收敛稳定, 初始学习率设置为 0.001, 并在每 20,000 个 epoch 后将学习率乘以 0.8; 优化器选用 Adam, 其参数设为 $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$ 。整个训练在搭载 NVIDIA GeForce GTX 1050 Ti 显卡的系统中耗时约 3 小时。

训练完成后, 模型可根据输入设备光电流 $\mathbf{I}$ 预测出入射光光谱 $\mathbf{P}$, 即

$$\hat{P} = \mathcal{N}_{\Theta_2}(F_{\Theta_1}(I_i)) \quad (3-8)$$

实验结果显示，该模型具有极高的预测效率：单个样本的预测时间约为 0.84 秒，并能在 1.04 秒内处理 10,000 个样本，得益于所采用的高性能并行计算平台。这种高效性为大规模实际应用提供了有力保障。

3.5 重建结果

3.5.1 峰谱重建结果

在本研究中，采用多样化方法评估所提重建算法的有效性。具体而言，利用四种不同类型的光谱——单峰激光光谱、双峰激光光谱、LED 光谱以及超连续光谱——照射光谱仪，随后进行光谱重建。每种光谱均具有独特特性和应用背景，能够充分检验方法的适应性和鲁棒性。

其中，单峰和双峰激光光谱均由 NKT Photonics 超连续激光器产生。NKT Photonics 作为全球领先的光电企业，专注于超连续激光器的研发与生产，其产品广泛应用于科研、医疗、通信、能源以及环保等领域。超连续激光器能够输出宽带、高稳定性的光谱，使其成为光谱测量技术中的重要工具。

为验证重建结果的准确性，采用商业光谱仪 YOKOGAWA AQ6374 测得的入射光谱作为真实值。该仪器具有高精度测量能力，为研究提供了可靠的参照。

为确保重建光谱的准确性，将重建结果与商业光谱仪测得的真实值进行对比。如图 3-10 所示，利用 NKT Photonics 超连续激光器在 480nm 至 820nm 范围内，生成了一系列窄带单峰光谱。单峰光谱的特征在于其只有一个波长展现出最高光强，而其它波长的光强显著较低，其带宽均为 10nm。

为使实验条件更贴近实际，通过调控激光器输出与光谱仪的接收参数，使得单峰光谱的强度与商业光谱仪获得的测量值基本一致。为便于展示和对比，所有光谱均进行了归一化处理，即将最大光强统一调整至 1，以消除数据尺度和量纲的影响，从而实现不同数据间的直观比较。

图中各单峰光谱的颜色与其对应波长一致，反映出单色光通常的颜色表现。重建光谱以实线展示，实线上标有采样点（采样周期为 0.44nm），而虚线则表示商业光谱仪测得的参考光谱。对比结果表明，方法能够高质量地重建 480nm 至 820nm 范围内的单峰光谱，验证了该重建策略的有效性。

如图 3-11 展示了单像素渐变带隙 PN 结光谱仪在 LED 光源下的性能。实验中，分别采用绿色、黄色和红色三种 LED 光源，其光谱中心分别位于约 530nm、620nm 和 590nm，带宽约为 30nm、10nm 和 70nm。利用所提方法对这些光谱进行重建，所得结果与商业光谱仪测得的参考数据非常接近。

在绿色 LED 光谱重建中，由于重建过程中存在强度与带宽混淆，重建带宽与参考值稍有差异；红色 LED 光谱的重建结果则与商业光谱仪吻合度极高；黄色 LED 光谱因带宽较大，重建时出现了轻微的光谱偏移，且重建带宽略大于参

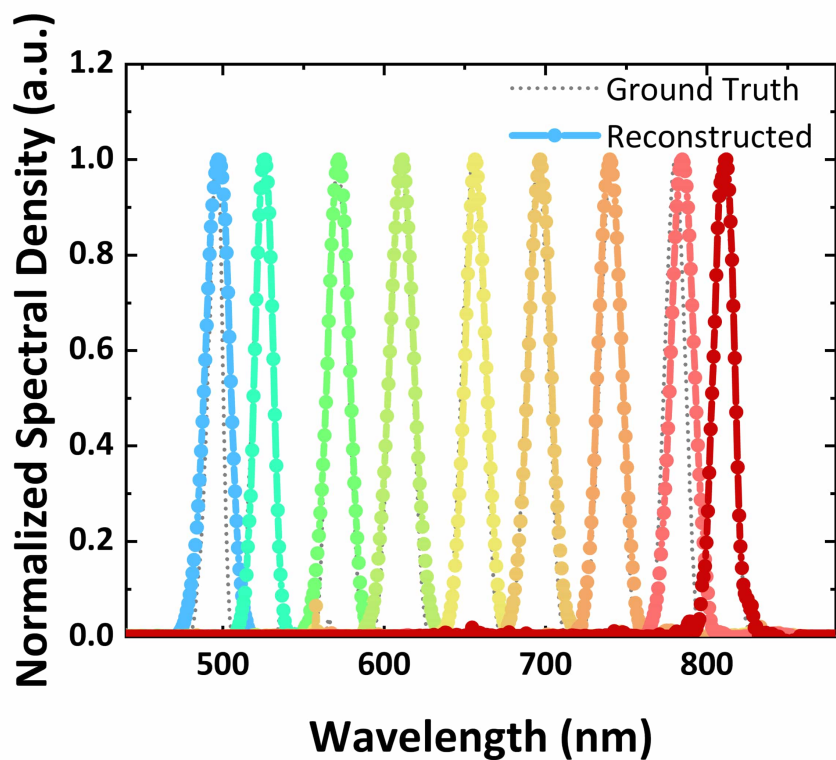


图 3-10 单峰光谱重建结果

Figure 3-10 Results of single peak spectral reconstruction

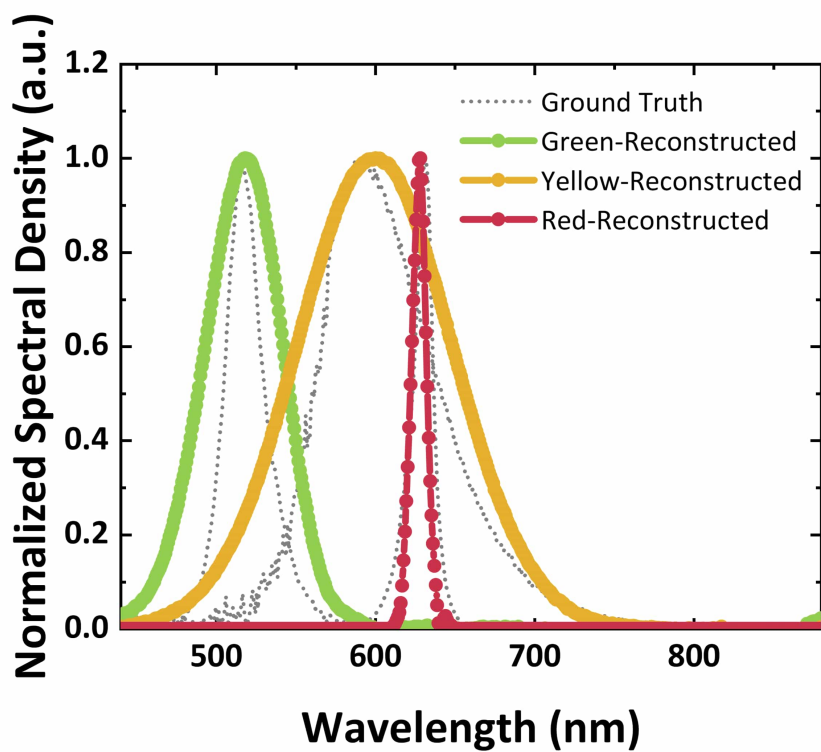


图 3-11 LED 光谱重建结果

Figure 3-11 Results of LED spectral reconstruction

考值。值得注意的是，尽管带宽增加导致绝对重建误差上升，但经过带宽归一化后的相对误差在各光谱条件下表现一致。这一结果进一步证明了本方法在面对不确定带宽的入射光时仍能实现准确重建，体现了其在实际应用中的广泛潜力。

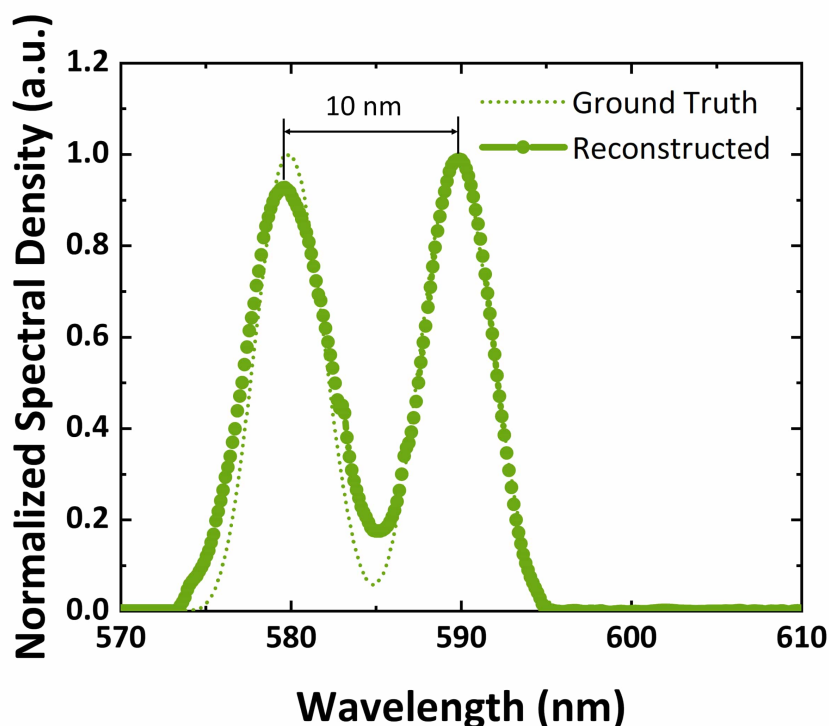


图 3-12 双峰光谱重建结果

Figure 3-12 Results of double-peak spectral reconstruction

本实验同时对光谱仪及所提方法的光谱分辨率进行了测试。光谱分辨率反映了仪器区分相近光谱峰的能力，通常通过间隔一定波长的双峰光谱评价，即能准确重建的最小峰间距离定义为仪器的光谱分辨率。如图 3-12所示，峰值波长分别为 580nm 和 590nm 的双峰光谱重建结果中，实线表示基于本方法恢复的入射光谱，虚线为商业光谱仪测得的真值。结果显示，在 590nm 峰值处，重建光谱几乎无波长偏移；而在 580nm 峰值处，虽有约 1nm 的偏差，但两个峰强度均与真值高度一致。该实验充分证明了本方法可以实现 10nm 的光谱分辨率，从而为其准确性和可靠性提供了有力支持。

为深入评估方法在波长分辨能力方面的表现，设计了密集采样实验，对间隔为 1 nm 的单峰入射光进行详细测量。实验选取 650 nm 至 660 nm 区间，生成一系列高精度单峰光谱，分别以黄色、褐色、深蓝色、灰蓝色、橙色、淡橙色、绿色、灰绿色、灰红色、玫红色、天蓝色和蓝色展示。在图 3-13中，实线表示本方法重建的光谱，虚线为商业光谱仪测得的参考数据。尽管各光谱峰值之间仅相差 1 nm，给波长分辨带来极大挑战，但实验结果显示，本方法能够以约 0.30 nm 的准确度精确重建各窄间隔峰值。图右上角进一步展示了重建峰值准确度与波长之间的关系，充分验证了方法在高分辨率波长采样下的优越性能。

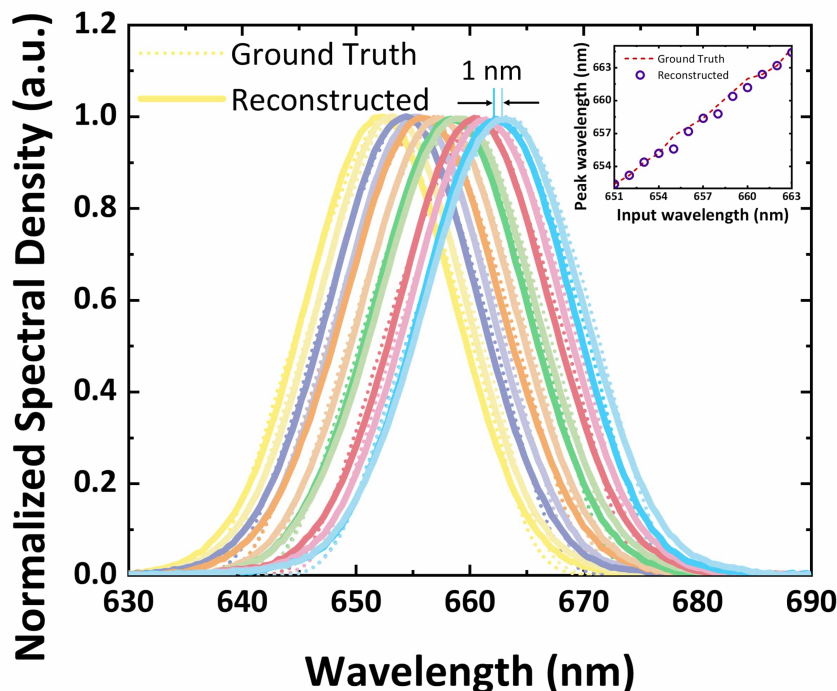


图 3-13 高精度光谱重建结果

Figure 3-13 Results of high-resolution spectral reconstruction

综上，实验结果充分证明了单像素渐变带隙 PN 结光谱仪在多种环境下实现高精度光谱重建的显著潜力。

3.5.2 光谱成像结果

为探究光谱仪在光谱成像中的应用，以绘画目标为实验对象，通过采集目标反射或透射的光谱信息来分析其色彩细节和光谱分布。如图 3-14所示，实验目标为一张印有“福”字的纸张，其中“福”字采用黄色颜料，背景则为红色，这种设计有助于评估光谱仪对不同颜色的辨识能力。

为保证均匀充足的照明，实验采用 LED 白光对场景进行照射，在透镜聚焦下获得清晰图像，从而精确捕捉目标的光谱信息。随后，在成像平面上放置直径为 $180\ \mu\text{m}$ 的探测器，其主要任务是接收并转换目标反射或透射的光信号。借助电动定位系统，探测器沿 X、Y 轴进行扫描以覆盖整个图像区域。

为了细致获取目标的光谱数据，图像平面被划分为多个采样单元，在每个单元内记录电压与光电流的变化，形成光电流曲线。X-Y 方向上的采样率设置为 100×100 ，确保了数据的充分性和高分辨率。通过这一实验设置，能够深入分析光谱仪在光谱成像中的性能与潜力。

图 3-15中展示了目标上每个像素均能对应得到的光电流曲线。图中以黄色和红色分别示出了“福”字颜料及背景颜料典型的光电流曲线，其中横轴表示电压，纵轴表示光电流强度，直观反映了电压变化对光电流的影响。随后，利用神经辐射场方法从光电流曲线中重建出相应的光谱曲线，这些曲线的横轴为波长，

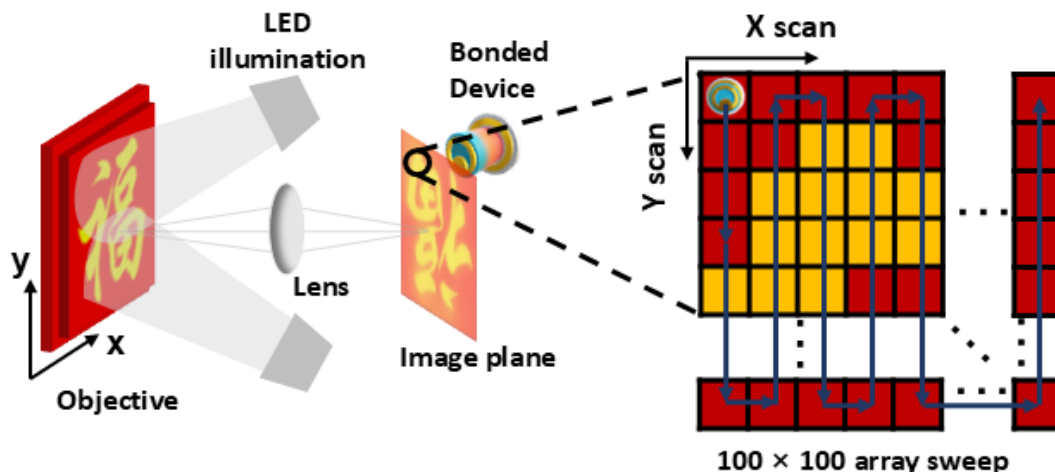


图 3-14 光谱重建装置

Figure 3-14 Spectral imaging setup

纵轴为光谱强度，分别以黄色和红色表示“福”字和背景区域重建得到的光谱。接着，根据每个像素的电流-电压数据重建出目标图像，并运用 CIE 1931 颜色空间映射转换为 RGB 图像，该步骤使得光谱图像转换为直观的彩色图像，便于后续分析。最终，重建的 RGB 图像在色彩和细节上与商业相机拍摄的图像基本一致，这验证了成像方法的准确性与可靠性。

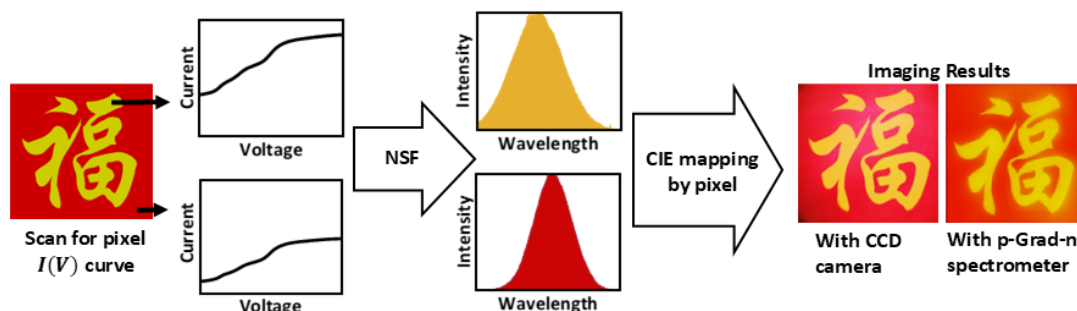


图 3-15 光谱重建流程

Figure 3-15 Spectral imaging setup and results

图 3-16中展示了通过实验重建获得的“福”字光谱图像，这些图像对应 480nm、510nm、540nm、570nm、600nm、630nm、660nm、690nm 和 720nm 九个不同波长通道。观察结果显示，在大多数波段下，“福”字与背景的轮廓均呈现鲜明分界，表明两者在光谱特性上存在显著差异，从而使得光谱成像能够有效地区分“福”字与背景。值得注意的是，在 600nm 波长下，“福”字与背景的边缘较为模糊。这是由于在该波段下，黄色与红色颜料反射的光谱极为相似，导致光谱成像难以实现准确区分。该现象与实验预期相符，进一步印证了实验方法的准确性与可靠性。

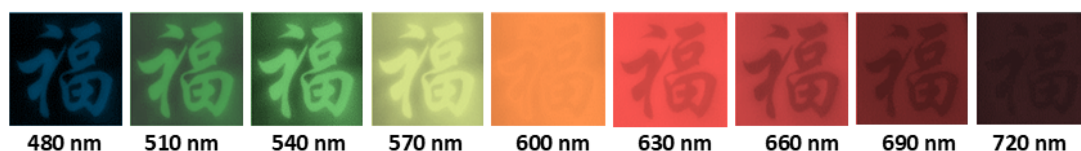


图 3-16 光谱重建结果

Figure 3-16 Spectral imaging results

3.5.3 测量误差的影响

为严格评估不同方法在测试阶段对噪声的鲁棒性，利用模拟数据生成了多种噪声水平的测试样本。具体来说，在仿真生成的测量 I-V 曲线中引入了均值为零、噪声幅度分别为测量最大值的 0.1%、1%、10%、20%、30%、50%、80% 和 100% 的高斯噪声。随后，分别采用了 Tikhonov 正则化方法和本文提出的方法对噪声数据进行重建测试。测试对象为峰值波长 700nm、带宽 10nm 的单峰光谱，其对应的光电流由仿真模型生成，并通过添加不同噪声强度来模拟实际情况。

图 3-17(a) 展示了 Tikhonov 正则化方法的重建结果。结果显示，在无噪声情况下，该方法能够准确还原光谱的峰值位置和强度；但当噪声仅为 0.1% 时，虽然峰的位置大致可以辨识，但其强度已出现明显偏差；而当噪声增加至 1% 时，重建结果几乎完全失效，光谱变为宽带且峰值模糊，难以估计准确的峰值波长，且强度与实际值存在较大误差。

相比之下，图 3-17(b) 展示了提出的方法在不同噪声水平下的重建结果。即使在加入 1% 噪声的情况下，该方法仍能准确重建出光谱的峰值波长与强度。即便噪声水平提升至 10% 时，入射光谱也能保持较好的重建效果。当噪声继续增加至 30% 时，重建峰值出现了轻微偏差；而在噪声达到 80% 时，峰值偏差进一步加大且偏差方向发生变化；在最大噪声水平 100% 下，神经网络方法仍可较好地估计出光谱的峰值位置和强度。

3.5.4 宽带重建

为深入探讨光谱仪对宽波段光谱重建的能力，对宽带单峰光谱进行了测试。实验以 700nm 为中心波长，选用带宽从 10nm 至 50nm（每 5nm 递增）的一组单峰光谱作为实验对象，这些光谱由 NKT 超连续激光器生成。光谱被引入后，光谱仪依次采集相应的光电流曲线。图 3-18 中展示了这九组光谱对应的光电流曲线，其中横轴为电压（负值代表负向偏压），纵轴为归一化后的电流强度。图例中，红色虚线表示实验测得的光电流，黄色虚线为理论预期的光电流，蓝色实线则为算法重建得到的光电流。

观察图中数据可以发现，不同带宽光谱产生的光电流曲线形状极为相似，这在一定程度上增加了光谱区分与重建的难度。同时，实测光电流、理论值与重建结果三者之间基本重合，表明重建算法与实验测量之间具有较高的一致性。这些结果为进一步探究光谱仪在宽带单峰光谱处理中的性能提供了依据。

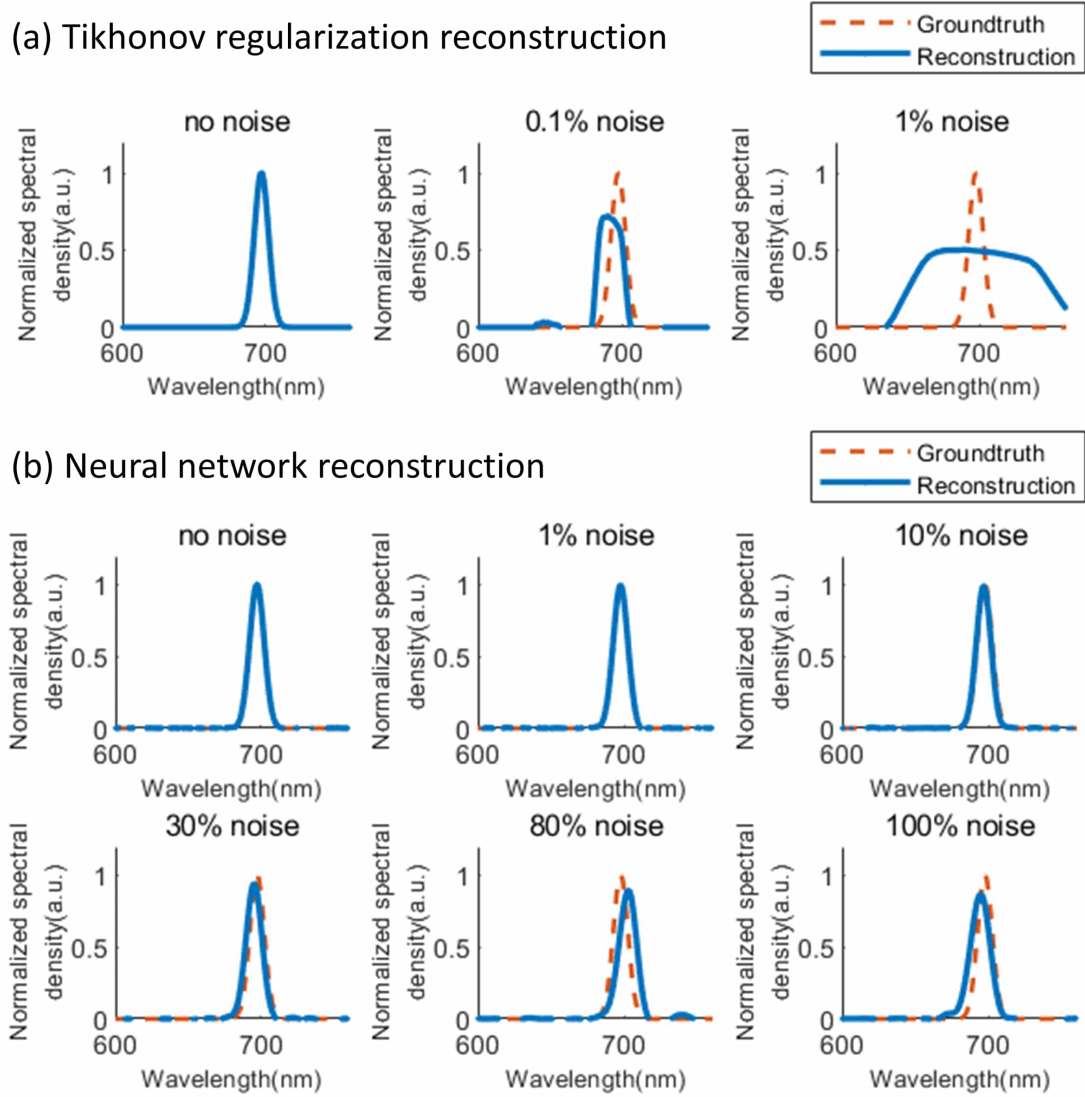


图 3-17 评估不同方法在测试阶段对噪声的鲁棒性

Figure 3-17 Evaluation of methods on noise.

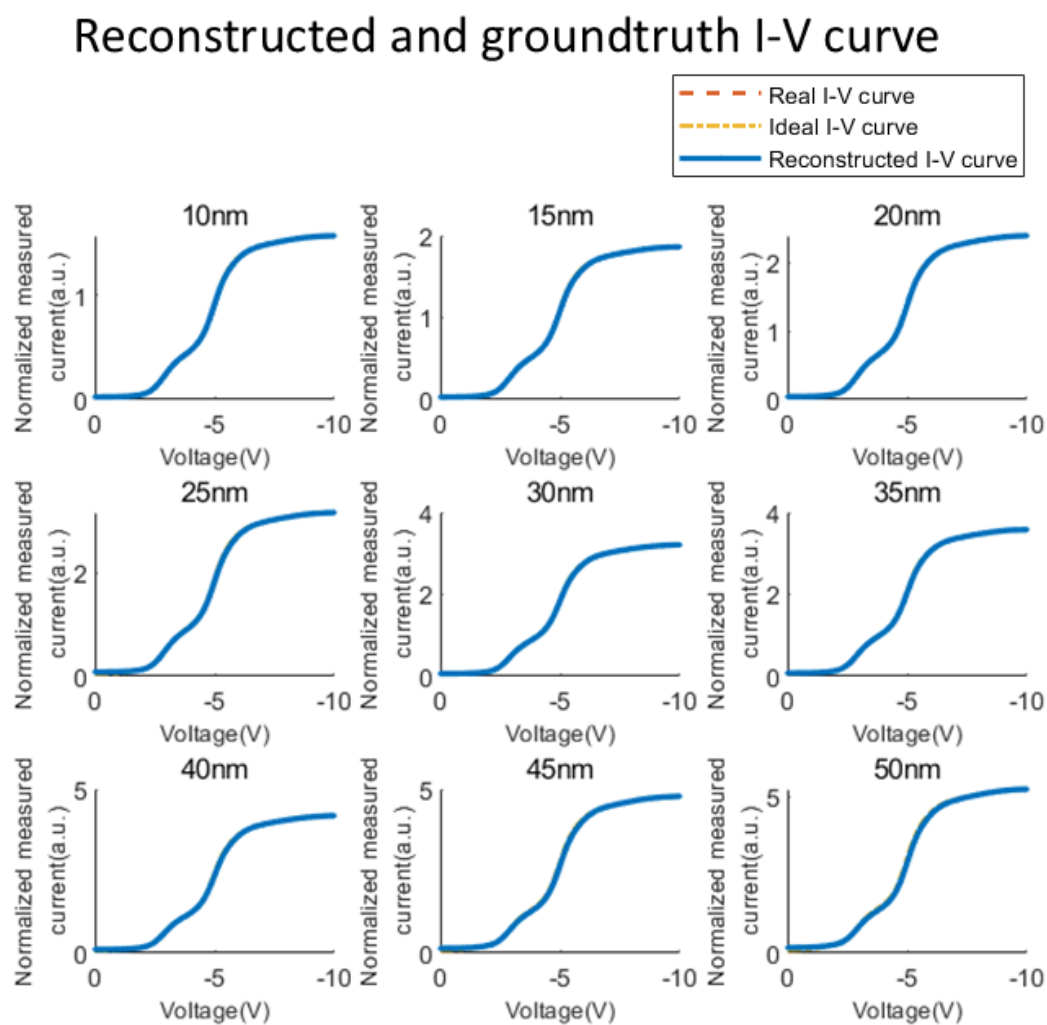


图 3-18 重建性能与峰宽之间关系的评估

Figure 3-18 Evaluation of performance on bandwidth.

图 3-19 中对图 3-18 的数据进行了局部放大，展示了负向偏压区间（-5.2V 至 -5.6V）内实测光电流、理论预期光电流和重建光电流之间的细微差异。为深入理解这些数据，从多角度分析了重建准确性，考虑了合成数据的不完美性以及算法重建误差。

首先，黄色的理论应得光电流曲线与红色的实测曲线之间存在的差异，主要归因于设备标定和实验测试误差。实验中获得 I-V 曲线并不能完美反映光谱仪内部响应函数与光谱间的乘积关系，这可能源于商用光谱仪内置探测器的固有响应与理论计算值之间的偏差。由此，训练所用的合成数据可能存在一定偏差，无法完全刻画真实数据分布，这反映了光谱仪在分辨率上的固有限制。为降低此类系统性误差，可能需要采用更精密的标定设备和更严格的测试流程。

其次，对红色实测曲线与蓝色重建曲线之间的差异进行了分析。理论上，理想重建应使二者完全重合，但由于理论与实测存在固有偏差，重建算法的目标在于修正这些偏差，而非直接拟合实测数据。为此，融合先验知识与神经网络算法成为必要手段，这有助于提高重建结果与真实光谱之间的一致性与可靠性。

最后，黄色理论曲线与蓝色重建曲线间的差异被视为评价重建方法性能的重要指标。因为理论光电流可通过仿真精确计算，重建结果越接近理论值，说明重建方法性能越高。与传统方法仅关注重现实测数据不同，提出的方法强调重建曲线与理论预期的一致性，从而为光谱重建提供了一种新的性能评价标准，并提高了整体准确性和可靠性。

如图 3-19 所示，随着入射光谱带宽的增大，重建误差呈逐步上升的趋势。这一现象可归因于以下原因：首先，带宽扩展会使光谱展现出更多非理想特性，增加了待处理信息量，从而对光谱重建构成更大挑战；其次，更宽的光谱通常包含更复杂的信息结构，要求重建算法不仅要精准还原峰值波长和强度，还需捕捉细微的边缘变化。

具体来看，当光谱带宽处于 10nm 至 15nm 时，重建结果非常准确；而在 20nm、25nm、30nm 和 35nm 范围内，虽然峰值波长与强度仍能准确恢复，但在边缘尾部（弱信号部分）出现了轻微偏差。当带宽进一步扩展至 40nm、45nm 和 50nm 时，重建结果在整体光谱形状上开始失真，尤其是边缘部分的偏差更为明显，尽管峰值参数依然准确。

需要指出的是，由于合成数据集中单峰光谱主要采用高斯函数进行建模，因而重建得到的 I-V 曲线与实际测量曲线之间存在一定差异，这在预期之内。总体而言，这些结果表明，光谱带宽的增加直接对应着重建难度和误差的上升，为实际应用中如何权衡光谱带宽与重建精度提供了参考。

图 3-21 中提供了分析不同误差类型之间关系的视角。根据数据可将误差划分为三类：1. 算法误差：指重建得到的 I-V 曲线与实际 I-V 曲线之间的平均绝对误差，反映了光谱重建算法本身的准确性和可靠性。2. 数据集误差：指理论应得光谱与实际测量光谱数据之间的平均绝对误差，这一差异可能源自实验设备误差、环境因素或数据处理过程的不完美性。3. 光谱误差：作为最终指标，光谱误

Reconstructed and groundtruth I-V curve in details

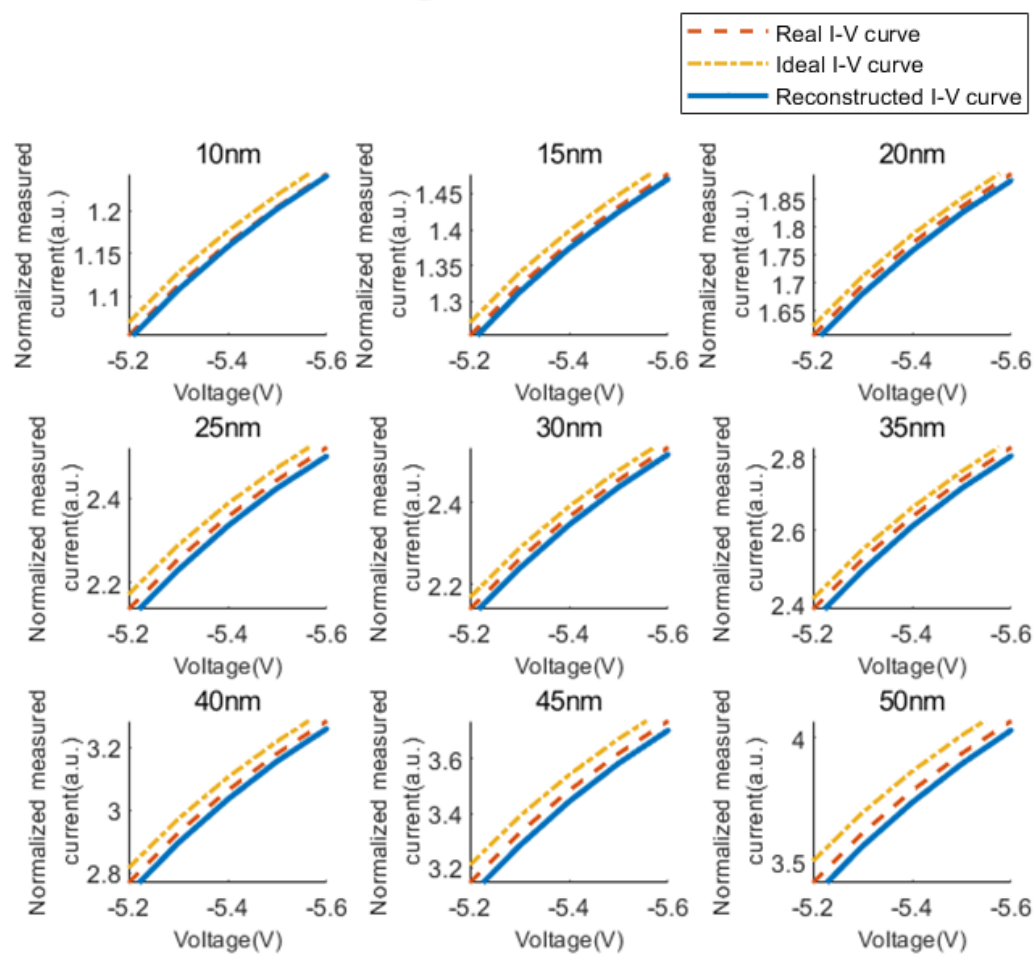


图 3-19 重建性能与峰宽之间关系的评估

Figure 3-19 Evaluation of performance on bandwidth.

Reconstructed and groundtruth spectrum

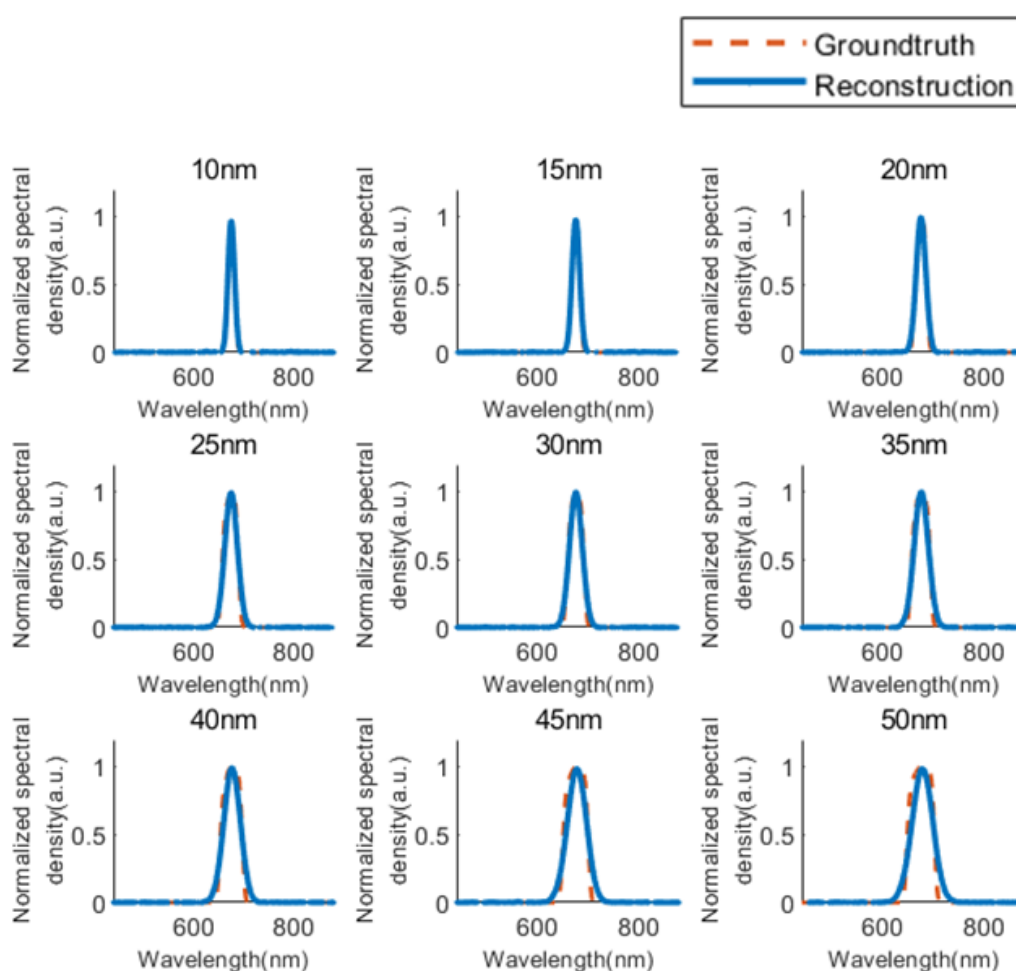


图 3-20 重建性能与峰宽之间关系的评估

Figure 3-20 Evaluation of performance on bandwidth.

差定义为重建光谱与真实光谱之间的平均绝对误差，综合反映了算法误差与数据集误差对最终重建结果的影响。

图中显示，随着数据集误差和算法误差的增大，光谱误差也呈现上升趋势。这表明，双方均会直接影响重建结果，因此在实际研究中，必须尽可能降低数据集误差与算法误差，以提高光谱重建的整体准确性。

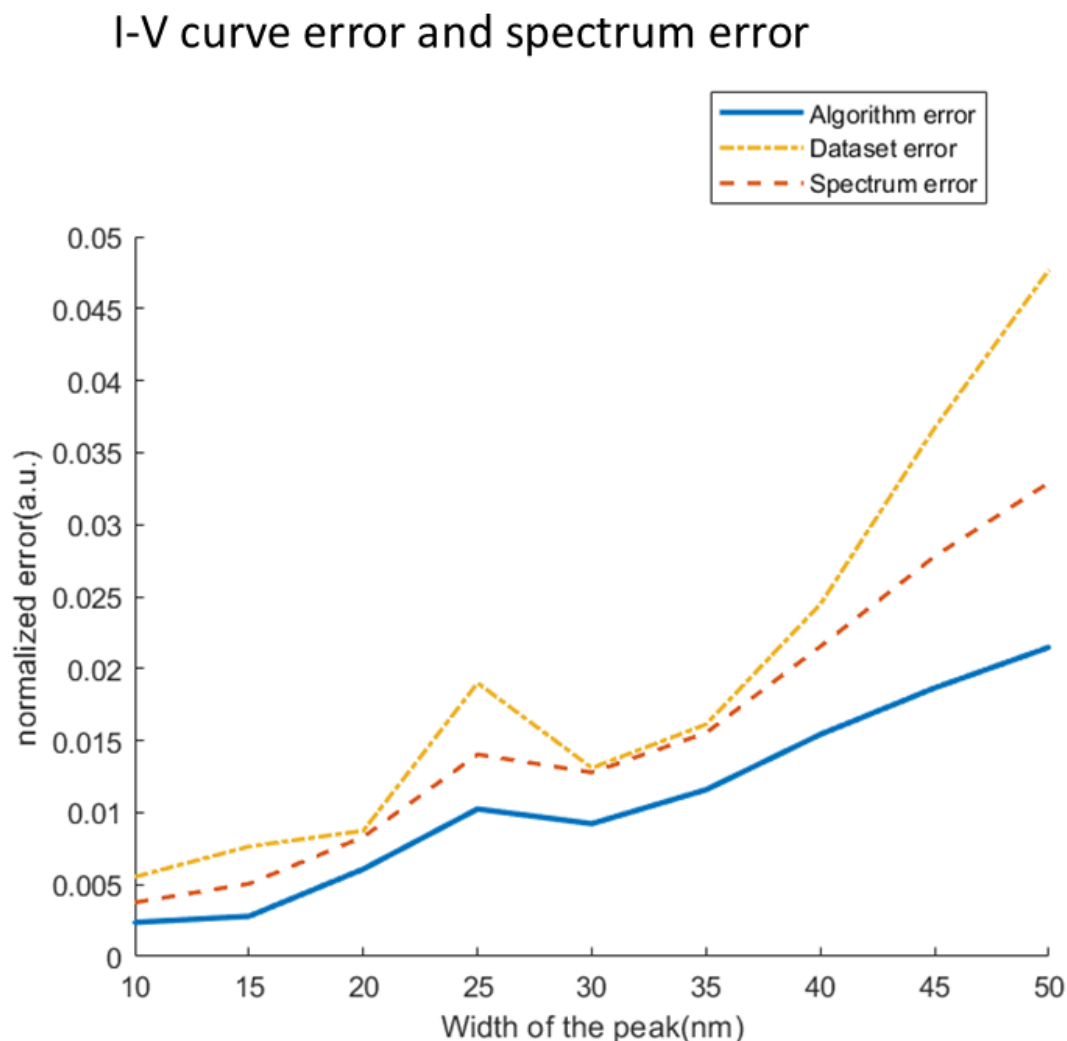


图 3-21 重建性能与峰宽之间关系的评估

Figure 3-21 Evaluation of performance on bandwidth.

3.5.5 多峰重建

为了更全面地评估光谱重建方法在处理复杂光谱情形下的有效性，采用合成数据生成并检测了具有三个和四个峰的人射光谱。首先构造三峰和四峰的合成数据，并以此训练神经网络模型。

如图 3-22所示，在三峰光谱重建任务中，神经网络方法成功重建出主要三个峰，重建的光谱和相应 I-V 曲线均与实际数据高度一致，尽管重建结果中存在

少量伪影。这些伪影可能源于光谱复杂性增加导致重建难度提升，或因训练数据与实际实验数据在分布上存在一定差异。此外，重建的 $I-V$ 曲线能够较好地捕捉到光谱仪模型的物理约束，但实际 $I-V$ 曲线与理想 $I-V$ 曲线之间仍存在一定偏差，这部分偏差可能与商业光谱仪固有的响应特性有关。

对于四峰光谱的重建，如图 3-23所示，神经网络方法依然能够重建出主要峰值，但整体重建准确性较三峰光谱略低。从图中可以看出，基于算法重建的 $I-V$ 曲线与实际测量曲线之间存在一定不匹配，反映出四峰光谱的复杂性对神经网络模型提出了更高要求。尽管引入的基于物理的细化策略在简单的平移和缩放操作上表现良好，但在处理复杂的四峰光谱时，仍出现一些伪影现象，这表明当前方法在适应实际 $I-V$ 曲线方面还有改进空间。

综上，实验结果表明，对于三峰和四峰的复杂光谱，神经网络方法均能够实现较为有效的重建，但在处理更高复杂度的光谱时，模型的表现有所下降。这一现象提示，在未来研究中需进一步优化方法，提高在复杂光谱重建中的准确性和可靠性。

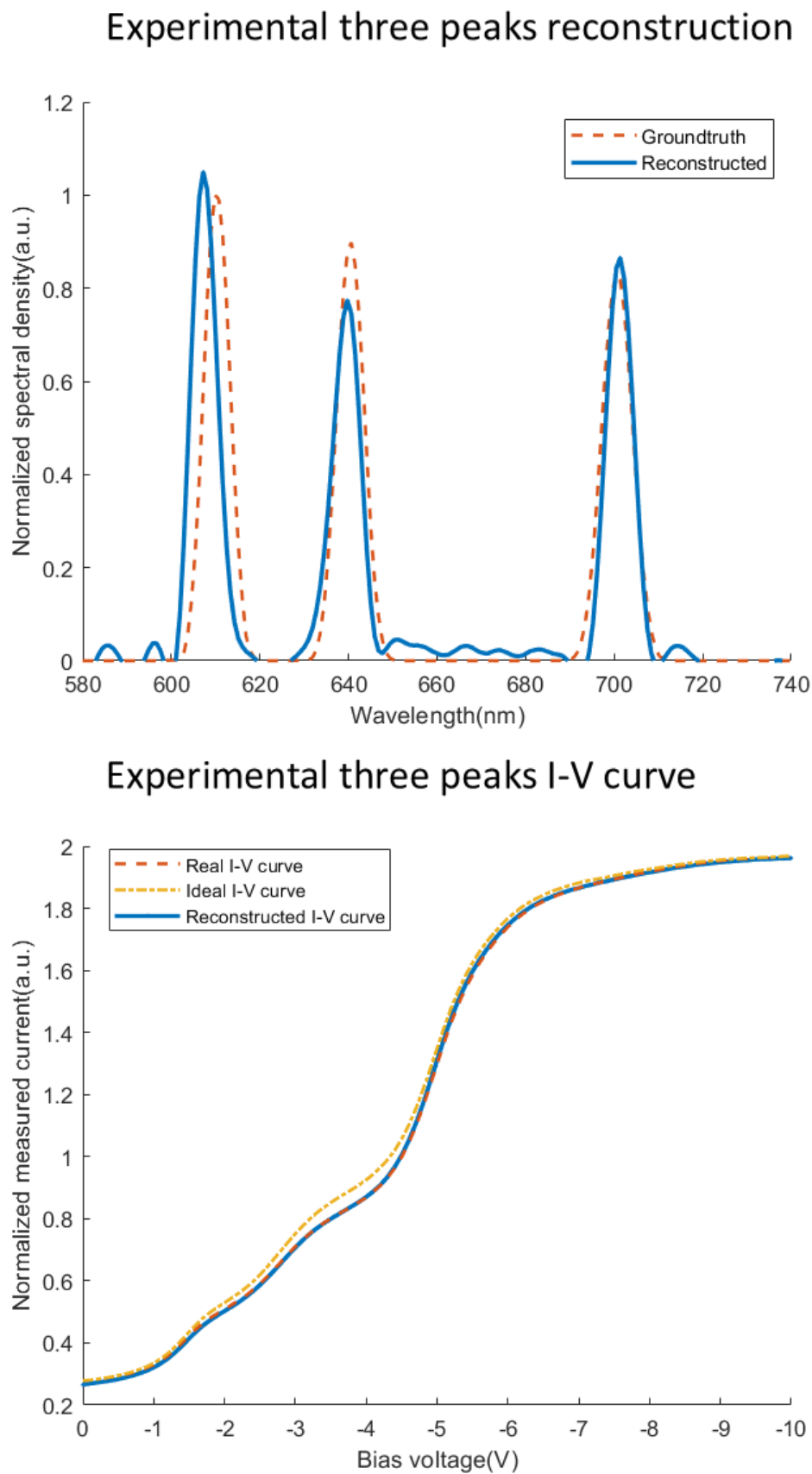
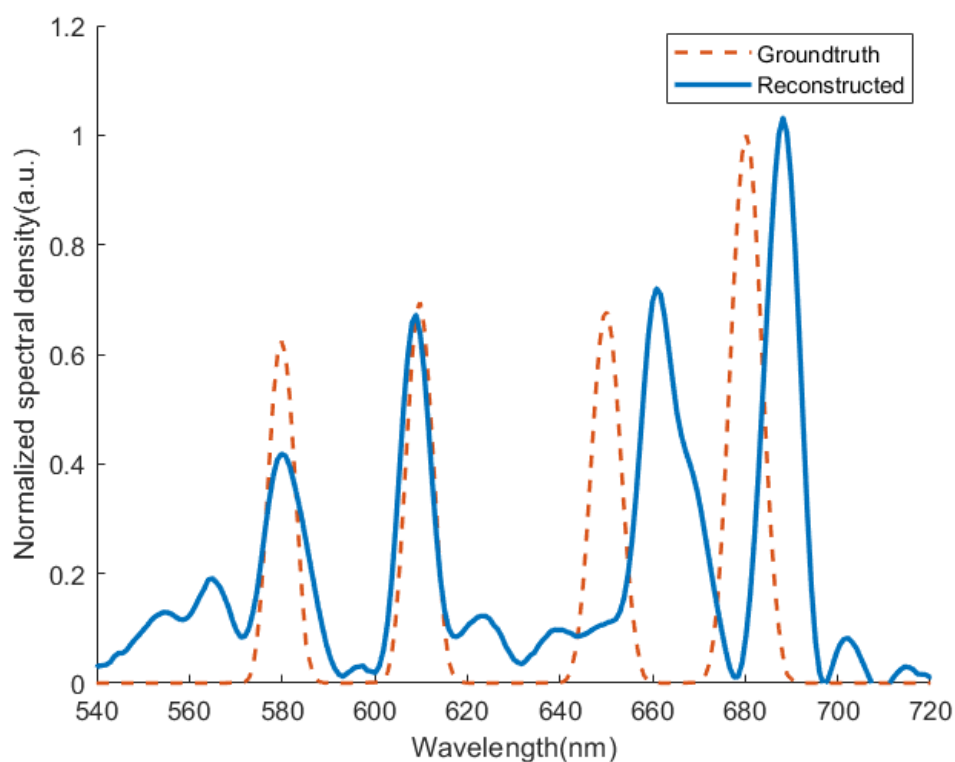


图 3-22 多峰光谱重建

Figure 3-22 Multi-peak spectrum reconstruction.

Experimental four peaks reconstruction



Experimental four peaks I-V curve

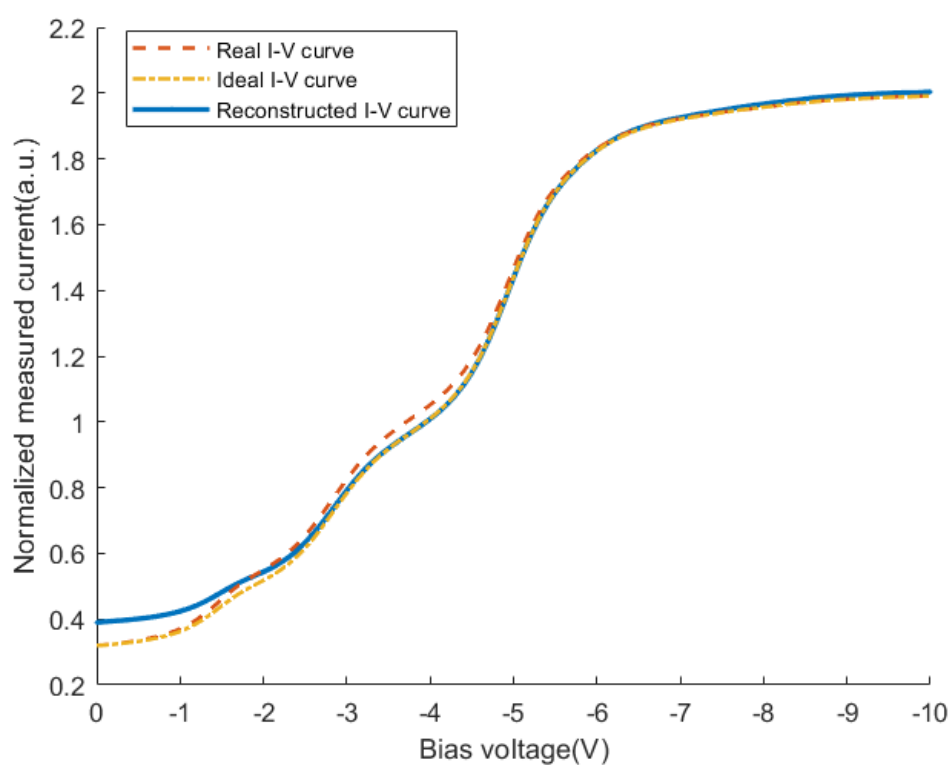


图 3-23 多峰光谱重建

Figure 3-23 Multi-peak spectrum reconstruction.

3.6 本章小结

光学光谱仪的小型化在便携式光谱检测、高光谱成像及可穿戴光谱分析等应用领域中具有重要意义。本研究基于渐变带隙 PN 结结构新型电压可调 III-V 族光谱仪，提出了一种基于神经光谱场的光谱重建方法，通过对累积电流进行逆向解析，实现了高精度、高鲁棒性的光谱重建。该研究为光学光谱仪的小型化发展作出了重要推进，拓展了其在诸多应用场景中的潜力。

未来研究将重点探索具有读出电路的单像素渐变带隙 PN 结光谱仪阵列的构建及其重建算法的研究，包括物理成像模型、神经网络架构，以实现实时原位光谱成像。这种“光谱相机”技术期望能够在无需色散光路或探测阵列的情况下，以超小尺寸实现多通道光谱信息获取。该技术有望突破集成化高光谱成像的技术瓶颈，在机器视觉、航天探测及大气探测等领域发挥重要作用。

第4章 基于神经瞬态场的自监督学习多光谱重建

4.1 引言

计算光谱成像在多个领域中都具有极为重要的作用。然而，这也伴随着一些困难和挑战。在实际的光学系统中，由于硬件系统的限制，光谱成像的前向模型通常呈现局部性，比如，探测器某个像素接收到的光学信号仅与其邻近像素有关。然而，对于一些特殊场景的光谱成像任务，例如非视距条件下的光谱成像，这一假设并不适用。因此，解决全局性前向测量模型的重建问题成为亟待解决的关键挑战。

非视距 (Non-line-of-Sight, NLOS) 成像旨在恢复相机视线之外的场景，例如，隐藏在墙后的物体。NLOS 的应用非常广泛，从遥感，到自动驾驶，再到危险环境中的救援任务。虽然早期的 NLOS 成像器主要受限于实验室环境，但最近的解决方案更多地利用通过如单光子探测器这样的可负担的时间解析探测器进行瞬态成像，作为商品捕捉系统越来越受欢迎。然而，到目前为止，几乎所有现有的基于瞬态的 NLOS 技术都是单色的。这主要归因于发射激光，其带宽非常窄。实际上，获取不同光谱的图像的能力对于推进 NLOS 在如检测和识别等高级计算机视觉任务中的使用至关重要。例如，很少有 NLOS 系统可以直接获取彩色图像，更不用说用于材料分析的不同颜色的对应反照率了。

为此，本文提出了一种在 NLOS 条件下生成彩色图像的多光谱神经瞬态场 (Multi-Spectral Neural Transient Fields, MSNeTF) 技术。不同于以往在各独立光谱上逐一进行 NLOS 重建并合成结果的方法，在谱间引入几何不变性约束，并开发了相应的神经重建框架。具体而言，构建了一种类似于视域成像中神经辐射场的光谱神经瞬态模型，并利用深度学习技术同时恢复场景的几何形状与光谱反照率。综合在合成与真实数据上的实验表明，MSNeTF 在视觉质量和重建精度上均优于当前最先进的单色 NLOS 方法。

值得注意的是，NLOS 光谱成像中的光谱反照率恢复问题与几何重建密切相关，两者通常被视为一个整体进行处理。现有方法多将 NLOS 场景表示为未知的三维反照率场和密度场，并通过测量瞬态图像进行反演求解；也有方法先参数化场景几何或反照率，再通过梯度下降等优化手段实现重建。有些启发式算法则借助瞬态数据的结构特征改进几何恢复。然而，专门针对光谱恢复并与几何信息结合的研究较少。最直观的方法在于单独对每个光谱进行 NLOS 重建后合成结果，但此策略忽略了跨谱几何一致性。虽然陈等人采用数据驱动的方法实现了 RGB NLOS 成像^[109]，但其研究主要依赖合成数据，并且在基于信号处理的系统中引入几何一致性先验仍存在较大挑战。

如图4-1所示，本文提出多光谱神经瞬态场实现非视距多光谱成像。图中展示了超连续可调谐激光器，该激光器能够自主发射不同波长的脉冲，而宽谱段探测器 SPAD 采集各波长下目标场景的光谱反射率。图中以红、绿、蓝三色分别展

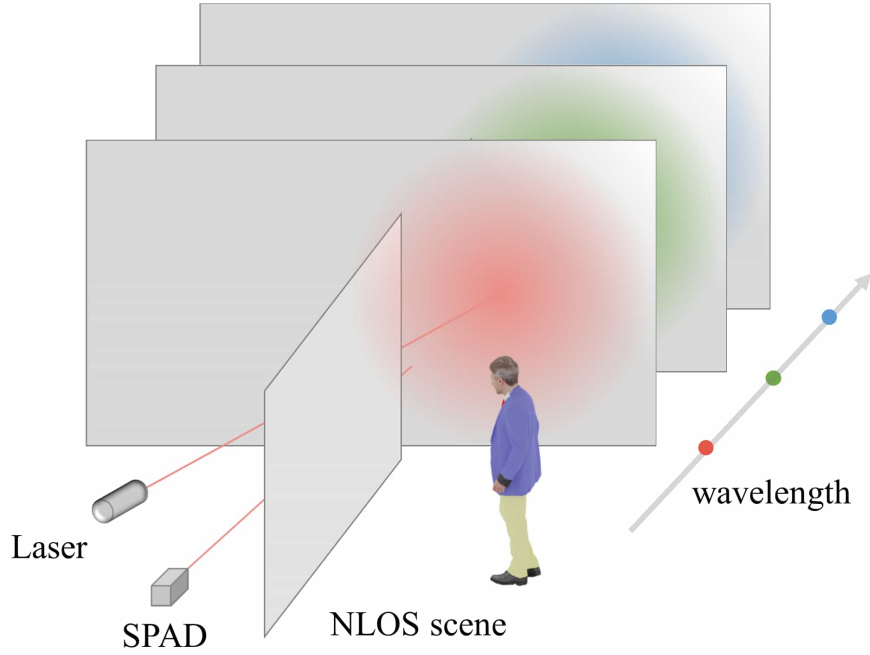


图 4-1 多光谱非视距成像

Figure 4-1 Multi-spectral non-line-of-sight imaging.

示了不同波长下的 NLOS 成像过程。实际操作中采用时分复用策略，依次以各波长扫描，进而获得完整的多光谱数据，从而恢复场景的几何结构与光谱反射率。MSNeTF 适用于通用非共焦设置，其中激光和 SPAD 可在中继墙上同时进行光栅扫描，通过在不同时间释放各波长脉冲捕捉光谱瞬态。

本文提出的深度学习方法在 NLOS 条件下生成彩色图像。数据采集系统采用宽光谱激光配合相应颜色滤镜。首先讨论了单波长 NLOS 重建逐通道实现多光谱成像的局限性，随后在神经重建框架中引入跨谱几何不变性约束，构建了一种新的光谱神经瞬态模型。该方法借鉴了视域下的神经辐射场模型与最新的 NLOS 神经瞬态场技术，采用专门的网络架构分别处理几何信息与波长依赖性反射率，并保持两者间的紧密联系。

此外，本文还提出基于泊松噪声模型的新型瞬态损失，相较于传统高斯模型具有明显优势，同时讨论了共焦与非共焦设置下的差异。综合合成数据和真实场景实验表明，MSNeTF 在视觉质量与重建精度上均优于现有最先进的单色 NLOS 方法，其获取的颜色信息为后续高级视觉任务提供了更充分的场景恢复数据。

4.2 系统分析

4.2.1 被动式与主动式

被动方法利用 NLOS 场景中的自然光照和现有相机恢复目标场景。传统相机（如 CCD 或 CMOS）具有多个通道（如 R、G、B）和高空间分辨率。例如，Bouman 等人利用角落遮挡物在不同视角下阻挡不同数量的光，定位并恢复二

维方向上移动的隐藏物体^[110]；Tancik 等人通过观察地板数据恢复室内隐藏部分^[111]；Saunders 等人在部分遮挡情况下生成隐藏物体的二维图像^[112]；Tanaka 等人利用入射光不同倾角分解成不同比例偏振光，结合偏振膜和常规相机恢复 NLOS 场景^[113]；Sharma 等人则采用深度学习对人数计数和人类行为分类进行研究^[114]。这类方法高度依赖于对场景先验信息（如几何形状、照明条件或遮挡结构）的恢复。

主动方法则结合专用光源（如投影仪或脉冲激光器）与探测器对 NLOS 场景进行重建。在时间解析 NLOS 成像中，Kirmani 等人首次采用瞬态成像相机和飞秒脉冲激光器恢复隐藏场景及实验室中不同颜色墙面的模拟效果，但因相机灵敏度和激光单色性限制，其重建精度和现实彩色实验效果存在不足。近期，一组 670 nm（红色）和 532 nm（绿色）的脉冲激光被用于捕捉 NLOS 场景的瞬态数据。Chen 等人进一步合成了包含 RGB 通道的大规模物体分类与识别的 NLOS 瞬态数据集^[109]，并在单波长激光采集的数据上应用了基于学习的方法。Musarra 等人则利用 490 nm、550 nm 和 610 nm 的超连续激光器重建了具有三种颜色的隐藏平面物体^[115]，但其工作未提供完备的理论前向模型。

受上述工作的启发，本文利用脉冲激光器配合可调滤光器（以特定波长为中心的带宽）和单像素或阵列格式的单光子探测器构建多光谱 NLOS 成像系统，同时推导出相应的物理前向模型。

4.2.2 共焦与非共焦

时间解析 NLOS 成像系统可分为三种设置：完全设置、单一设置和共焦设置。完全设置中，照明和检测点均发生变化；单一设置中，照明或检测点保持固定，而另一者变化；共焦设置中，照明与检测点重合。前两种统称为非共焦设置，其中照明和检测点分别作为椭球两个焦点，而共焦设置则构成非共焦设置的三维子集。

Ahn 等人利用两对光圈控制激光扫描，并采用单个 SPAD 构建了完全设置，但所采数据较难处理^[116]。Buttafava 和 Liu 分别构建了单一设置—SPAD 聚焦中继墙中心，而脉冲激光束扫描墙面^[117,118]。此外，还可利用 1D 条纹相机或 SPAD 的一维/二维阵列，同时将激光束定向至中继墙的固定点来构建单一设置。条纹相机虽能提供高达 2 ps 的时间分辨率，但价格昂贵且校准困难；而 SPAD 近年来成为了一种较经济且便捷的选择。

作为首个共焦设置，O'Toole 等人提出了光锥变换方法^[119]，通过光束分裂器使单像素 SPAD 与脉冲激光器同轴。他们利用更大功率的激光器改进了 NLOS 成像系统，并建立了大规模 NLOS 场景数据集，同时采用费马路径对场景进行重建。共焦设置在校准和数据采集上更为简便，但当检测与照明点完全重合时，中继墙的首次反射信号可能会占主导。事实上，非共焦设置的重建算法也适用于共焦设置；通过将共焦测量近似为照明和检测点的中点，可实现两种设置之间的转换。

本文提出的 MSNeTF 的前向模型适用于这两种设置，并且提出了两种多光谱瞬态测量方案：一种结合单个 SPAD 或 SPAD 阵列与带可调滤光器的超连续脉冲激光器；另一种利用与多光谱滤光器耦合的 SPAD 阵列配合超连续脉冲激光器。

4.3 非视距成像的传统方法

传统算法（如光锥变换）仅恢复隐藏物体的反射率或反照率，忽略了 NLOS 成像中光的波长信息。为了重建具有颜色属性的 NLOS 场景，一个直观的方法是假设不同波长下的 NLOS 场景在成像过程中是独立的；但在存在如色散等复杂物理效应的情况下，这一假设便不成立。针对这些场景，逐个波段使用传统算法恢复隐藏物体的反射率或反照率存在两个问题：首先，隐藏物体在各个波长下的几何形状是相同的，而仅恢复每个波段的反射率会忽略几何形状信息，导致深度估计和遮挡检测不准确；其次，传统算法依赖于在中继墙上均匀扫描得到的瞬态数据，在逐个波段进行 NLOS 重建时要求每个波段的扫描模式完全一致，而光谱 NLOS 成像中对每个波段进行全面数据采集既耗时又不必要。

将 NLOS 场景表示为一个神经网络。输入到神经网络的瞬态被分为两部分：NLOS 场景中一个点的位置 $\mathbf{x}$ ，以及该点相对于中继墙上的照明和检测点的仰角和方位角 $(\alpha_i, \beta_i, \alpha_d, \beta_d)$ 。在 NLOS 设置中，MLP 的输出是 NLOS 场景的概率或密度， $\sigma(\mathbf{x})$ ，以及反射率， $\rho(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d)$ 。将 MLP 的结构和参数 W 视为 NLOS 场景的参数化表示。

基于光传播的物理原理，推导出一个可微分的 NLOS 前向模型，用于预测 NLOS 场景的瞬态。接着，将 NLOS 重建问题表示为最小化预测瞬态与测量瞬态残差的优化问题，从而获得 MLP 的最优参数。优化完成后，再通过体积渲染生成 NLOS 场景的彩色图像。流程图如图4-2所示。

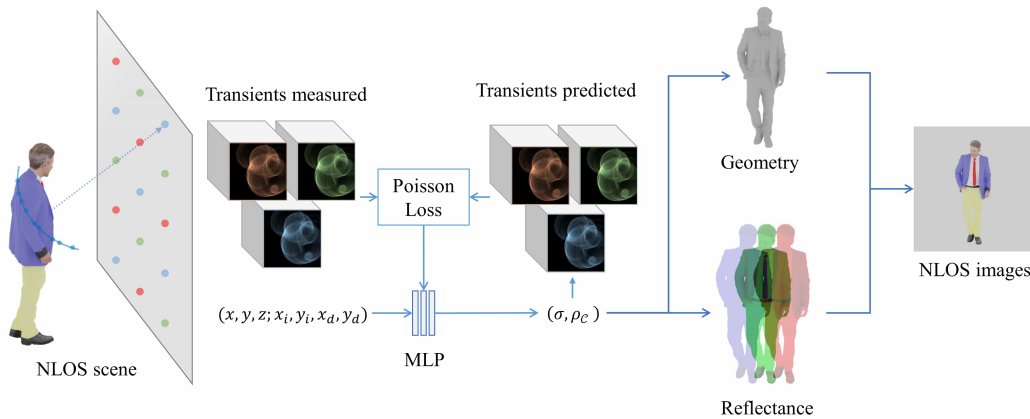


图 4-2 提出的 MSNeTF 流程

Figure 4-2 Pipeline of our proposed MSNeTF.

提出的 MSNeTF 流程如图4-2所示，以一位穿着蓝色西服和淡黄色裤子的人作为隐藏物体示例。在重建流程中，首先根据墙面上预设的扫描点，在物体可能

存在的空间中进行球形采样。接着，用一个神经网络隐式表达未知的非视距场景，因此采样得到的三维点坐标及光线方向可以输入网络，形成对应的物体表示。MLP 的输出包括密度和每通道的反射率，并依据光谱非视距成像模型，在每个球面上积分生成预测瞬态。当预测瞬态与实测瞬态差异最小时，即认为神经网络表达的物体信息与实际物体一致。

为最小化预测瞬态与实测瞬态间的差异，一个以神经网络参数为自变量的优化问题被求解，通常使用 L2 范数作为差异度量。同时，本文引入泊松损失来优化 MLP，以同时生成物体几何和光谱反射率。在迭代优化过程中，随着损失函数逐步下降，当其变化趋于平稳时，便认为神经网络训练完成。

训练完成后，对物体可能存在的空间进行均匀采样，生成物体几何和光谱反射率的存在可能性网格。一方面，利用图形学方法将网格转换为三角微元，实现后期渲染和展示；另一方面，将光谱反射率转化为光谱贴图并直接映射到三角微元上，最终渲染出完整的物体图像。

4.3.1 多光谱神经瞬态场

考虑在非视距 (NLOS) 成像中的波长 λ :

$$\tau(x_i, y_i, x_d, y_d, t, \lambda) = L(\lambda)S(\lambda)\Gamma_0 \iint_{E(x_i, y_i, x_d, y_d; ct)} \frac{1}{r^4} \sigma(\mathbf{x}) \rho(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d, \lambda) \quad (4-1)$$

其中 $L(\lambda)$ 是激光的发射光谱函数， $S(\lambda)$ 是单光子雪崩二极管 (SPAD) 的检测效率光谱函数， Γ_0 是系统常数。 $E(x_i, y_i, x_d, y_d; Ct)$ 是一个半椭球体，以 (x_i, y_i) 和 (x_d, y_d) 为焦点， CT 为光路。 $\sigma(\mathbf{x})$ 是非视距场景中的点 $\mathbf{x}$ 的密度，这是与波长无关的。 $\rho(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d, \lambda)$ 是点 $(\mathbf{x})$ 的反射率。 (α_i, β_i) 和 (α_d, β_d) 显示了在不同入射和出射方向上反射率的几何效应。非视距场景的反射率在不同的波长 λ 下是变化的。 Ψ 是从墙 (x_l, y_l) 到达非视距场景并反射回墙 (x_w, y_w) 的光线路径。 dA 是微面的面积。Eqn. (4-1) 给出了光谱的连续描述。但在多光谱非视距成像中，由于系统带宽的限制，测量瞬态仅覆盖有限的波长范围。在许多情况下，只需要特定通道内的光谱反射率，而非连续的光谱数据。因此，本文推导了适用于有限波长范围的非视距成像模型。

假设滤波器包含不同的通道 $C = \mathcal{R}, \mathcal{G}, \mathcal{B} \dots$ ，将在通道 C 下测量的瞬态表示为 τ_C ，对应的工作波长为 $[\lambda_{Cmin}, \lambda_{Cmax}]$ 。为了简化这个表达式，首先将非视距场景表示为 $(\sigma, \rho(\lambda))$ 并定义 $\mathcal{R}$ 为一个与激光发射光谱 $L(\lambda)$ 和探测器响应光谱 $S(\lambda)$ 都无关的变换：

$$\mathcal{R}(\sigma, \rho(\lambda)) = \iint_{E(x_i, y_i, x_d, y_d; ct)} \frac{1}{r^4} \sigma(\mathbf{x}) \rho(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d, \lambda) \exp \left(- \int_{\Psi} \sigma ds \right) dA \quad (4-2)$$

然后将 Eqn. (4-1) 表示为：

$$\tau(x_i, y_i, x_d, y_d, t, \lambda) = \Gamma_0 L(\lambda) S(\lambda) \mathcal{R}(\sigma, \rho(\lambda)) \quad (4-3)$$

由于系统具有带宽，很难在特定波长上测量瞬态。因此，定义通道的 C 瞬态为：

$$\tau_C(x_i, y_i, x_d, y_d, t) = \int_{\lambda_{Cmin}}^{\lambda_{Cmax}} \Gamma_0 L(\lambda) S(\lambda) \mathcal{R}(\sigma, \rho(\lambda)) d\lambda \quad (4-4)$$

只考虑非视距场景在通道 C 的平均反射率：

$$\rho_C(r, \theta, \phi) = \int_{\lambda_{Cmin}}^{\lambda_{Cmax}} \rho(r, \theta, \phi, \lambda) d\lambda \quad (4-5)$$

试图从在通道 C 测量的瞬态， $\tau_C(x_i, y_i, x_d, y_d, t)$ ，重构反射率 ρ_C 。然后在 Eqn. (4-4) 中应用积分的平均值定理：

$$\tau_C(x_l, y_l, x_w, y_w, t) = \Gamma_0 L(\lambda_{\xi}) S(\lambda_{\xi}) \int_{\lambda_{Cmin}}^{\lambda_{Cmax}} \mathcal{R}(\sigma, \rho(\lambda)) d\lambda \quad (4-6)$$

注意，变换 $\mathcal{R}$ 是关于波长 λ 的线性的；因此可以将积分放在 ρ 上，并得到：

$$\tau_C(x_l, y_l, x_w, y_w, t) = \Gamma_0 L(\lambda_{\xi_C}) S(\lambda_{\xi_C}) \mathcal{R}(\sigma, \rho_C) \quad (4-7)$$

整个非视距成像模型由将变换 $\mathcal{R}$ 代入 Eqn. (4-7) 得出：

$$\tau_C(x_i, y_i, x_d, y_d, t) = \Gamma_0 L(\lambda_{\xi_C}) S(\lambda_{\xi_C}) \iint_{E(x_i, y_i, x_d, y_d; ct)} \frac{1}{r^4} \sigma(\mathbf{x}) \rho_C(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d) \quad (4-8)$$

给定激光发射光谱曲线 $L(\lambda)$ 和 SPAD 响应光谱曲线 $S(\lambda)$ ，将可以从采集的瞬态 $\tau(x_i, y_i, x_d, y_d, t)_C$ 重构非视距场景在通道 C 的密度 σ 和反射率 ρ_C 。

4.3.2 共焦与非共焦实施

提出的方法在非共焦设置中定义，但在实际应用中，全面采样中继墙上所有照明点与检测点的配对在不同通道下既数据量巨大又冗余。因此，针对共焦和非共焦设置分别提出了两种多光谱非视距成像方案。

如图4-3(b)所示，共焦设置集成了超连续可调谐激光器、可调谐滤波器、宽谱单光子探测器和中继墙面等关键组件。所用超连续激光器可自主调节波长，而宽谱的 SPAD 使得在不同波长激光照射下得以准确测量目标场景的光谱反射率。为直观展示成像过程，图中以红、绿、蓝三色分别呈现了三种波长激光的非视距

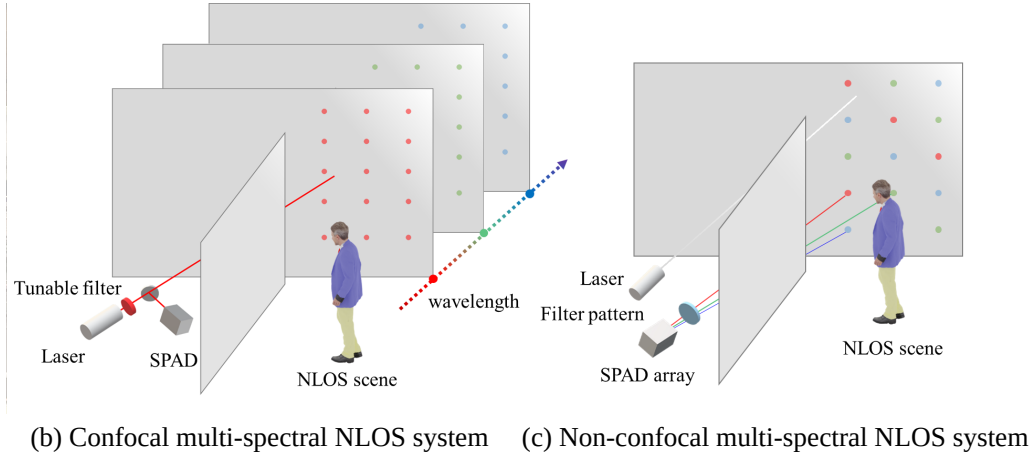


图 4-3 共焦与非共焦的系统设置

Figure 4-3 Setup of confocal and nonconfocal setup.

成像。实际操作中，通过时分复用方案在不同时刻扫描不同波长，有效支持了目标场景几何结构与光谱反射率的重建。

如图4-3(c)所示，另一种非共焦设置同样包括超连续激光器和中继墙面，但采用宽带单光子探测器阵列或单光子相机，取代了共焦设置中的单像素探测器。在此配置下，超连续激光器直接发出宽带激光，无需借助可调谐滤波器生成窄带激光；而探测器阵列利用各像素的分光功能，可在单次曝光中同时获取各波长下的目标信息，从而极大提高成像速度，达到每秒多帧实时成像的潜力，相比之下，共焦设置的激光单点扫描耗时较长。

提出的 MSNeTF 方法即基于这一通用非共焦设置，其关键在于激光与 SPAD 同步在中继墙上进行光栅扫描。操作过程中，激光在各不同时刻发射不同波长的脉冲，以捕获目标的光谱瞬态特性。

4.3.3 共焦系统

传统的共焦非视距成像系统使用分束器使激光和单光子雪崩二极管 (SPAD) 在中继墙上的同一扫描点上重合。通常，激光在固定波长下操作。使用一个复合激光与一个可调滤波器来实现多光谱非视距成像，如图 4-3 所示。复合激光产生不同波长的光混合。滤波器可以控制不同波长的光的传输特性。在共焦设置中，多光谱非视距成像可以通过在不同时间控制滤波器的中心波长并测量非视距场景在不同波长范围的反射率来实现。方法支持大多数基于体积的重建算法，例如，LCT^[119]，F-K^[120]，和 PF^[121]。在这种情况下，照明点 (x_i, y_i) 和检测点 (x_d, y_d) 在相同的位置。等式 (4-8) 可以简化为：

$$\tau_C(x_i, y_i, t) = \Gamma_0 L(\lambda_{\xi_C}) S(\lambda_{\xi_C}) \iint_{E(x_i, y_i; ct)} \frac{1}{r^4} \sigma(\mathbf{x}) \rho_C(\mathbf{x}, \alpha_i, \beta_i) \quad (4-9)$$

4.3.4 非共焦系统

SPAD 阵列的开发使用户能够进行无扫描非视距成像。在共焦设置中，非视距成像系统需要照射中继墙并检测不同点以获取非视距重建的足够信息。在使用 SPAD 阵列的非共焦设置中，激光只需要照射墙上的固定点，SPAD 阵列同时收集中继墙上许多检测点的瞬态。为了实现非共焦多光谱非视距成像，将 SPAD 阵列的每个像素涂覆不同中心波长的滤波器，或者在 SPAD 前加装滤波器阵列。在这种设置中，激光照射相同的点 (x_i, y_i) ，例如， $(0, 0)$ 。等式 (4-8) 可以简化为：

$$\tau_C(x_d, y_d, t) = \Gamma_0 L(\lambda_{\xi_C}) S(\lambda_{\xi_C}) \iint_{E(0,0,x_d,y_d;ct)} \frac{1}{r^4} \sigma(\mathbf{x}) \rho_C(\mathbf{x}, \alpha_i, \beta_i, \alpha_d, \beta_d) \quad (4-10)$$

4.4 噪声模型和损失函数

4.4.1 单光子探测器噪声模型

如图 4-4 所示，MSNeTF 方法在 4.3.1 中对无噪声和干扰的模拟数据表现出了良好的性能。在非视距成像中，由于光需要经过三次弹射反射，才能从激光器传到 SPAD，因此，得到的信号强度极其微弱。在对 MSNeTF 方法在实际非视距成像系统中测得的噪声瞬态对重建误差的影响进行分析时，发现噪声是主要的影响因素。

为了提高重建质量，将 SPAD 基非视距成像系统的噪声模型整合到了 MSNeTF 前向模型中。这个整合的模型如下所示：

$$\mu_C(x_i, y_i, x_d, y_d, t) = P(N \tau_C(x_i, y_i, x_d, y_d, t) * j(t) + N \tau_b(x_d, y_d)) \quad (4-11)$$

在这个模型中， τ_C 表示的是在通道 C 中，非视距场景的瞬态。在模型中， τ_b 表示的是在环境光下生成的瞬态。 N 表示的是激光的脉冲数， $j(t)$ 表示的是整个系统的时间抖动，而 d 是暗计数率。 P 是泊松分布。 $\mu_C \in \mathbb{Z}^+$ 是 SPAD 对时间实例 t 的光子计数。

理论上，由于脉冲数 N 和时间抖动 $j(t)$ 对不同的照明点和检测点是不变的，可以从系统校准中获得这两者。然而，背景在不同的检测点上变化。为了模拟这种变化，使用了一个定制的多层感知机 (MLP)。这个 MLP 的输入是检测点 (x_d, y_d) ，输出是对应检测点的背景瞬态 τ_b 。在重建过程中，将优化背景的 MLP 与非视距重建同步进行，以提高重建效果。

如图 4-4 所示，第一行依次展示了在不同环境光建模条件下的 MSNeTF 重建结果，分别为未考虑、已知以及估计环境光的情况；第二行则对应展示了非视距场景的真实值、估计环境光和实地测量的结果。真实场景中，以斯坦福大学经典“Stanford Bunny”为例，各部位分别涂有独具光谱特性的颜色，如黄色尾部、紫

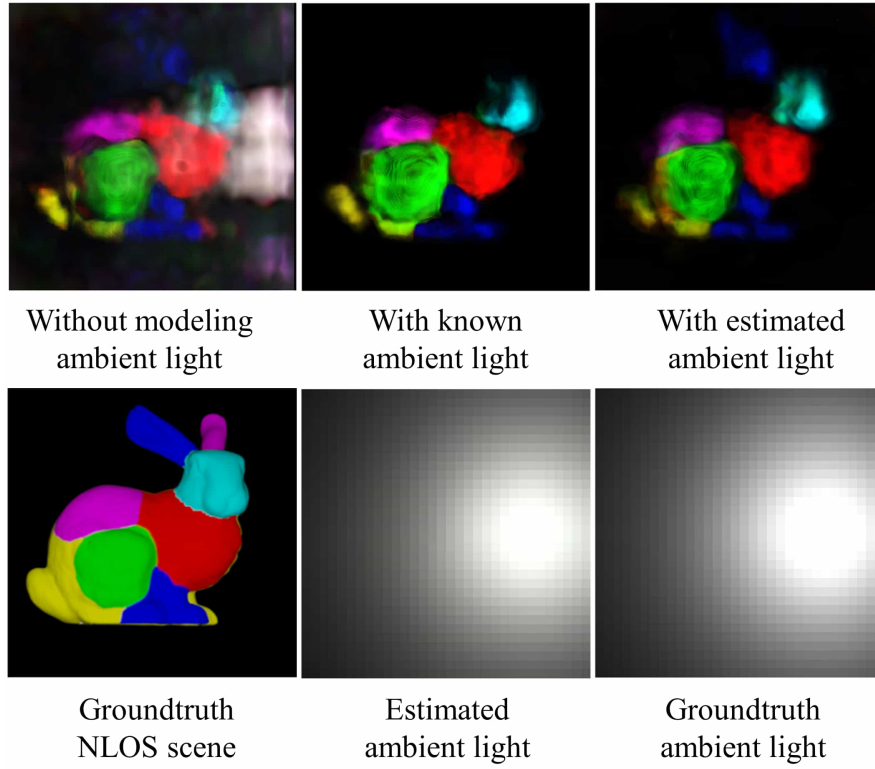


图 4-4 带环境光建模的 MSNeTF 结果

Figure 4-4 MSNeTF with ambient light modeling.

色背部、绿色大腿、蓝黄相间的前腿、红色脖颈、青色面部及蓝紫色耳朵。这些颜色变化为检验重建算法的光谱和几何重建能力提供了参照。

在未考虑环境光建模时，重建结果出现明显误差：兔子尾部的正确黄色中混杂了橙色和红色，前腿颜色也偏差明显；此外，背景光照引起的噪声将空白区域误判为实体，形成伪影。采用已知背景光强度与分布信息进行重建后，伪影得以消除，但仍无法准确复现部分细节，如蓝色耳朵，主要因直流成分虽被去除，但泊松噪声引起的信号方差仍然存在。

在 MSNeTF 中引入神经网络对环境光建模后，不仅实现了对物体几何和光谱的准确重建（尤其在兔子面部），同时还成功估计出背景光信息，与实测数据吻合。这一结果对自然环境下的非视距成像具有重要意义，通过定制的环境光估计 MLP，MSNeTF 显著提升了整体性能并实现了环境光的准确定量。

4.4.2 基于模型的预测和泊松损失

MSNeTF 的前向模型在实验设计中发挥关键作用，其可微性显著简化了设计流程。基于此特性，创新地通过分解泊松分布，向系统中引入了噪声模型，从而提升了整体模型的鲁棒性。

使用 L2 范数对神经网络进行优化的过程可以表示为：

$$\mathcal{L} = \|\mu_C(x_i, y_i, x_d, y_d, t) - \mu_{C_m}(x_i, y_i, x_d, y_d, t)\|_2 \quad (4-12)$$

在这个公式中， μ_C 表示根据公式 (4-8) 和公式 (4-11) 预测的非视距场景的瞬态， μ_{Cm} 表示测量的瞬态。在这种优化操作中并未考虑泊松分布，因此使用 L2 范数的重建质量表现较差，如图 4-5 所示。后续采用了最大似然估计方法对泊松瞬态进行处理，具体过程如下：

$$\mathcal{L} = \sum_{x', y', t, C} (\mu_C - \mu_{Cm} \log(\mu_C)) \quad (4-13)$$

在这个过程中，在重建过程中寻找 MLP 的最优参数 w ，目标是使得预测 μ_C 的概率 $P(\mu_C)$ 最大，从而使损失 $\mathcal{L}$ 最小：

$$\min_w \sum_{x', y', t, C} (\mu_C - \mu_{Cm} \log(\mu_C)) \quad (4-14)$$

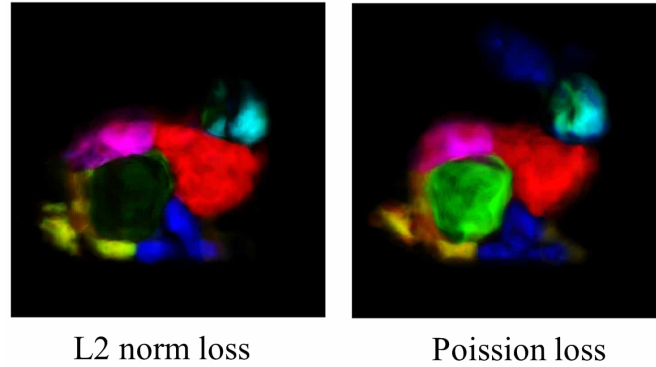


图 4-5 泊松损失函数的结果

Figure 4-5 Results with Poisson loss.

图4-5展示了使用 L2 范数损失与泊松损失在单光子成像重建中的视觉效果对比。首先，基于 L2 范数的重建结果存在明显问题：兔子大腿上的绿色未被准确复现，而出现大面积黑色覆盖，高强度的泊松噪声造成信号均值偏移；兔子尾部颜色斑驳，蓝色耳朵完全缺失，这反映出 L2 范数对泊松噪声敏感性不足。

相对而言，采用泊松损失函数后的重建效果显著提升。绿色大腿恢复完整；兔子青色面部细节丰富，如眼部凹陷得到了明显呈现；耳朵部分也得到了完整重建。由此可见，在单光子成像重建问题中，泊松损失函数更能有效应对噪声干扰，提升重建精度。

4.4.3 泊松损失

在本研究中采用最大似然估计处理泊松瞬态，可从观测数据中推导出最可能产生这些数据的参数值。

为便于描述，用向量 μ_C 表示所有预测瞬态的集合，其中每个瞬态 $\mu_C(x_i, y_i, x_d, y_d, t)$ 均以标量 μ_C 表示，描述在时间 t 、位置 (x_i, y_i, x_d, y_d) 及颜色 C 下的瞬态值。

进一步地,引入了衡量预测瞬态 μ_C 与相应测量值 μ_{C_m} 一致性的概率,这一关键参数直接决定了预测结果的准确性。该概率可通过以下公式计算:

$$P(\mu_C) = \frac{\exp(-\mu_C)\mu_C^{\mu_{C_m}}}{(\mu_{C_m})!} \quad (4-15)$$

在这个公式中, $\exp(-\mu_C)$ 是泊松分布的一个典型形式,描述了在给定的时间和空间条件下,发生某种事件的概率。而 $\mu_C^{\mu_{C_m}}$ 和 $(\mu_{C_m})!$ 则分别代表了预测的瞬态的测量值的指数和阶乘。

在实验中,测量每个瞬态 μ_C 的过程是独立的,所以所有预测的瞬态 μ_C 与对应的测量值 μ_{C_m} 一致的概率可以通过下面的公式进行计算:

$$P(\mu_C) = \prod_{x_i, y_i, x_d, y_d, t, C} \frac{\exp(-\mu_C)\mu_C^{\mu_{C_m}}}{(\mu_{C_m})!} \quad (4-16)$$

在这个公式中, $\prod_{x_i, y_i, x_d, y_d, t, C}$ 表示了对所有的时间、空间和颜色条件进行的乘积操作。

接下来,将概率 $P(\mu_C)$ 的负对数似然表示为 F :

$$F = -\log(P(\mu_C)) = \sum_{x_i, y_i, x_d, y_d, t, C} \log(\mu_{C_m}!) + \sum_{x_i, y_i, x_d, y_d, t, C} (\mu_C - \mu_{C_m} \log(\mu_C)) \quad (4-17)$$

在这个公式中, $\log(\mu_{C_m}!)$ 是预测的瞬态的测量值的对数阶乘,而 $(\mu_C - \mu_{C_m} \log(\mu_C))$ 则表示预测的瞬态与其对应的测量值的差值的对数。

最后,定义泊松损失 $\mathcal{L}$ 为:

$$\mathcal{L} = \sum_{x_i, y_i, x_d, y_d, t, C} (\mu_C - \mu_{C_m} \log(\mu_C)) \quad (4-18)$$

4.5 实验结果

对合成和实际瞬态进行实验,定性和定量地验证 MSNeTF 模型,并与最先进的方法进行比较,包括 LCT^[119], F-K^[120], 和 PF^[121]。

4.5.1 合成数据生成

在本研究中,开发了一种非视距渲染器,用以生成难以直接观测的三类非视距场景的瞬态数据。为实现这一目标,采用基于物理的合成方法,并借助渲染引擎 Mitsuba2 生成强度图与深度图的配对数据,这些数据采自中继墙上预定的不同位置,进而用于构建非视距瞬态。

所有非视距场景均设定为弥散光传播模式，即光线在传播过程中发生散射，进而影响观测结果。模拟了三种场景：九种不同颜色的补丁、一只斯坦福兔子和两个人，分别位于中继墙约 0.5 米、0.5 米、0.8 米及 1.2 米处。

在实验设计中，考虑了共焦与非共焦两种设置。在共焦配置下，瞬态的空间分辨率设为 32×32 ，时间分辨率为 4ps，同时墙体尺寸为 1 米 $\times$ 1 米；而在非共焦设置中，为获得更清晰的图像，瞬态空间分辨率提升至 64×64 ，并将墙体宽度扩展至共焦配置的两倍，以更好地匹配非共焦视野要求。

4.5.2 实际数据测量

图 4-6(a) 展示了设计制作的实际系统实物照片。系统中，超连续激光器发出的宽谱激光经过可调滤波器调控，转换为窄带激光脉冲。滤波器的高精度设计确保了激光脉冲的稳定性与高能量。随后，窄带激光脉冲经分束器分离为两路，并由振镜反射至中继墙面实现预定反射。在反射过程中，信号以球面波形式触及被测物体，并再次反射回中继墙面。最终，经过振镜和单光子探测器接收的反射信号转化为脉冲，后者被送至计数器进行精确记录。

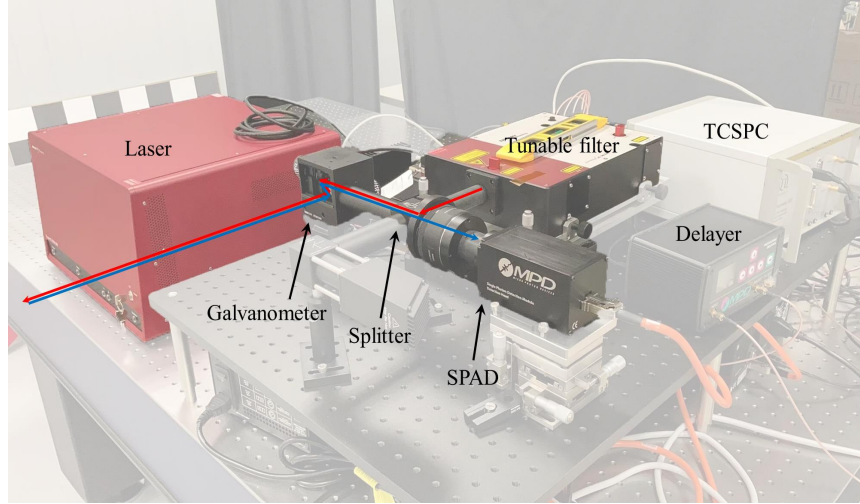
使用 MPD 快门 SPAD、Canon EF 50 mm f/1.8 消色差镜头以及 PicoHarp 300 构建检测系统。激光器 (SuperK EXTREME/FIANIUM, NKT Photonics) 具备 410–2400 nm 连续光谱，并通过配套的可调滤波器选取特定波长，以满足实验要求。

在共焦配置中，经偏振光束分离器 (Thorlabs VA5-PBS251) 准直后的光束，由 2D 扫描镜 (Thorlabs GVS012) 引导。利用延迟器 (PicoQuant PSD-065-A-MOD) 门控直达光，SPAD 记录由中继墙 (白板) 反射回来的非视距场景光子。

系统安装于距离中继墙 1.0 米处，扫描区域为 $0.7 \text{ m} \times 0.7 \text{ m}$ ，整体时间抖动约 100 ps。实验对象为穿有绿色人造丝内衣及红色棉外衣的人体模型。RGB 数据采集采用重复率 39 MHz、平均功率 100 mW 激光，并在 510 nm、590 nm 和 670 nm 三个带宽内分别进行测量。测量瞬态数据具有 64×64 空间分辨率及 4 ps 时间分辨率。

4.5.3 单光子模型的验证

图 4-4 直观展示了前向模型引入噪声模型后对非视距瞬态重建性能的提升。直接使用 MSNeTF 成像会产生伪像，但在充分考虑环境光噪声后，MSNeTF 可实现高质量场景重建。此外，在未知环境光情况下，MSNeTF 不仅重建效果与知情条件相当，还能精确估计环境光。具体而言，在共焦设置下，对比结果显示，由于空间分辨率受限，LCT 和相位场方法呈现模糊效果；F-K 方法虽恢复部分细节，但未能完整捕捉场景轮廓。而 MSNeTF 不仅恢复了颜色和轮廓，还细致呈现了场景细节。在非共焦设置下，传统方法均丢失了非视距场景中的关键细节，人物信息恢复困难；而 MSNeTF 则成功重建补丁、兔子，并恢复了人物的颜色和轮廓，充分证明了其优越性。



(a) Multi-spectral NLOS imaging system

图 4-6 硬件系统

Figure 4-6 Hardware system.

4.5.4 定性比较

如图 4-7(a) 展示了共焦设置下 MSNeTF 与 SOTA 方法的重建对比。LCT 和 PF 方法能较好恢复非视距场景的基本轮廓，但在细节复现上存在不足；F-K 方法在细节恢复上表现较好，但其反照率重建中噪声较多。相比之下，MSNeTF 不仅准确还原了场景的颜色与轮廓，其在几何细节上（如兔耳及书本等局部结构）也表现出色。

图 4-7(b) 展示了非共焦设置下的对比结果。利用^[120]提出的非共焦近似技术，将瞬态数据在共焦与非共焦设置间进行转换。传统方法（LCT、F-K 及 PF）仅能恢复场景的主要部分，且在处理两人场景时因人物距离中继墙较远而出现失败现象。而 MSNeTF 在非共焦条件下能够有效提取场景颜色与细节，在两人远距离案例中较好地恢复了人物轮廓和位置，尽管部分细节仍有所缺失。

如图 4-8 展示了多光谱非视距成像系统测得的瞬态数据经 MSNeTF 重建后的结果，并与 SOTA 方法进行了对比。尽管所有四种方法在恢复人体模型夹克的几何结构与颜色方面均未达到理想效果，但 MSNeTF 与 PF 在躯干部分的还原更为逼真，相较之下，LCT 与 F-K 表现欠佳。这一结果验证了 MSNeTF 在实现精确彩色非视距重建方面的潜力。在本研究部分对真实人体模型的瞬态数据进行了比较实验，采用 SOTA 方法和 MSNeTF 两种策略。在 670 nm（红色，见中行）和 550 nm（绿色，见下行）波长下捕获单通道瞬态，然后合成彩色图像（上行）。

如图 4-9 对真实的 RBbar 瞬态数据进行了详细对比，采用 SOTA 方法与 MSNeTF 算法。在 670 nm 和 500 nm 波长下分别采集了对应红色（中行）和蓝色（下行）的单通道瞬态数据，随后融合生成彩色图像（上行）。

如图 4-10 以真实的零食瞬态为对象，比较分析了 SOTA 方法与 MSNeTF 的

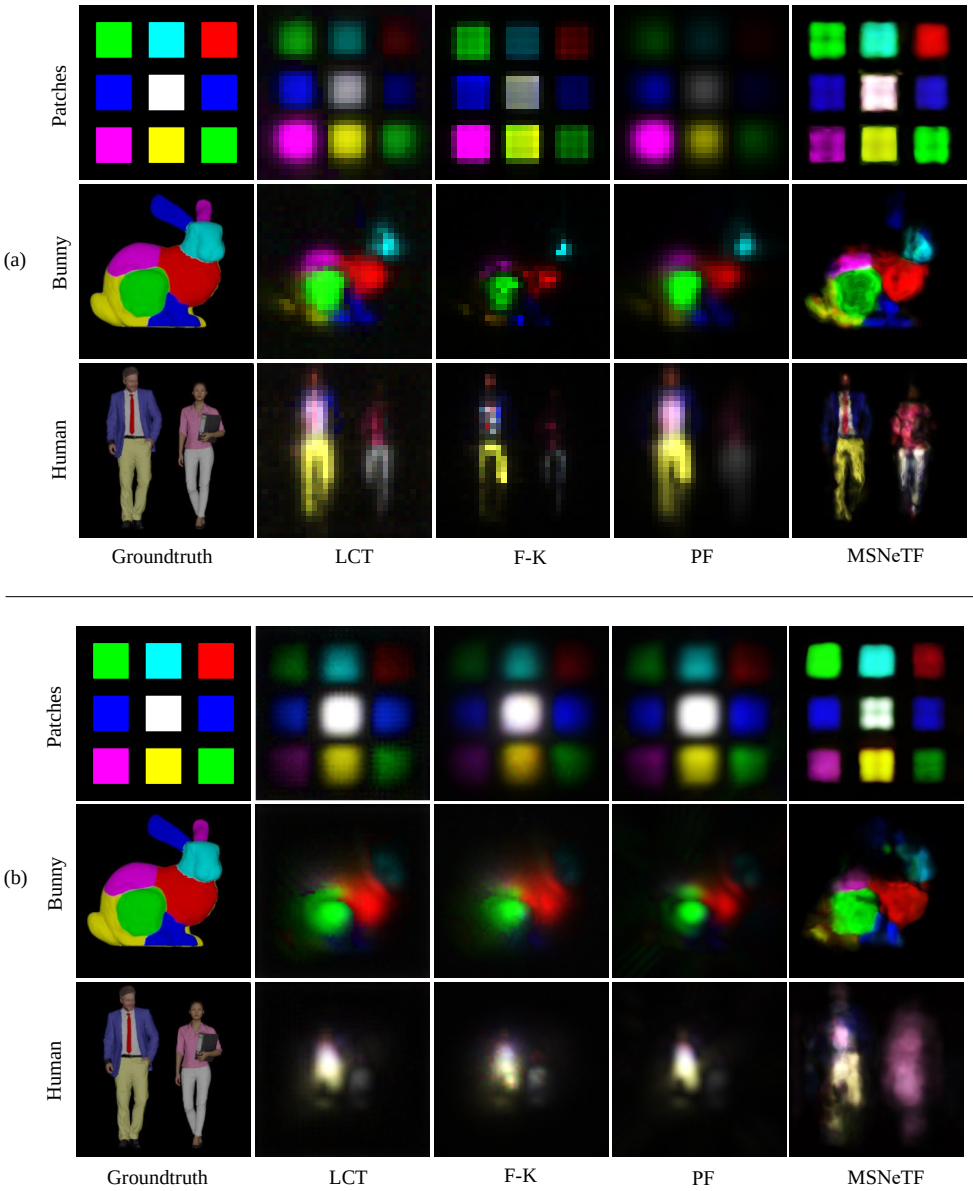


图 4-7 共焦和非共焦场景重建结果

Figure 4-7 Results on confocal and non-confocal setup.

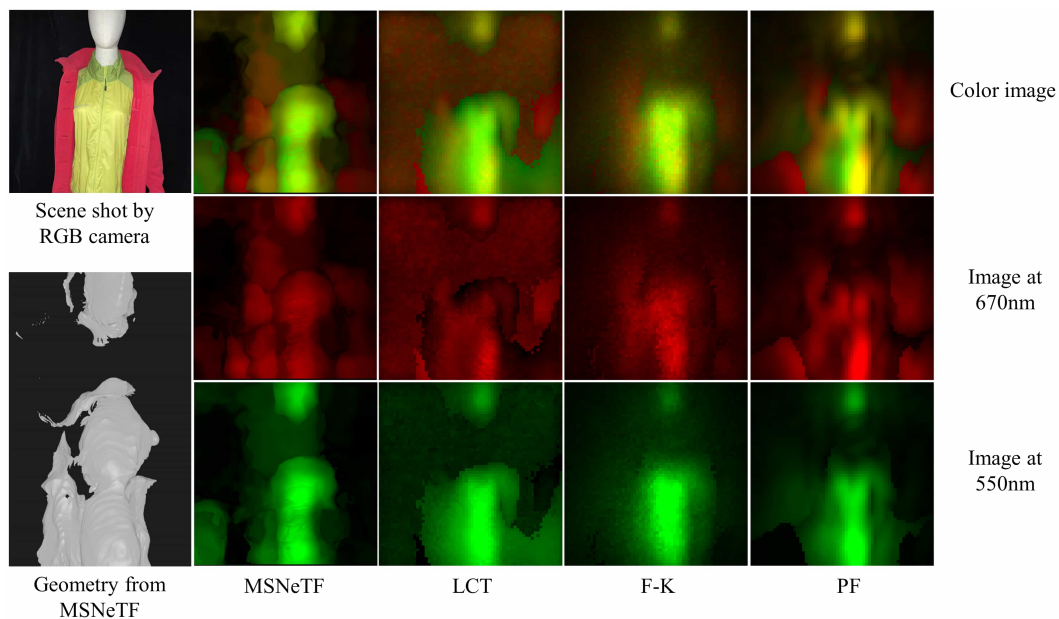


图 4-8 人体模型上的结果
Figure 4-8 Results on human model.

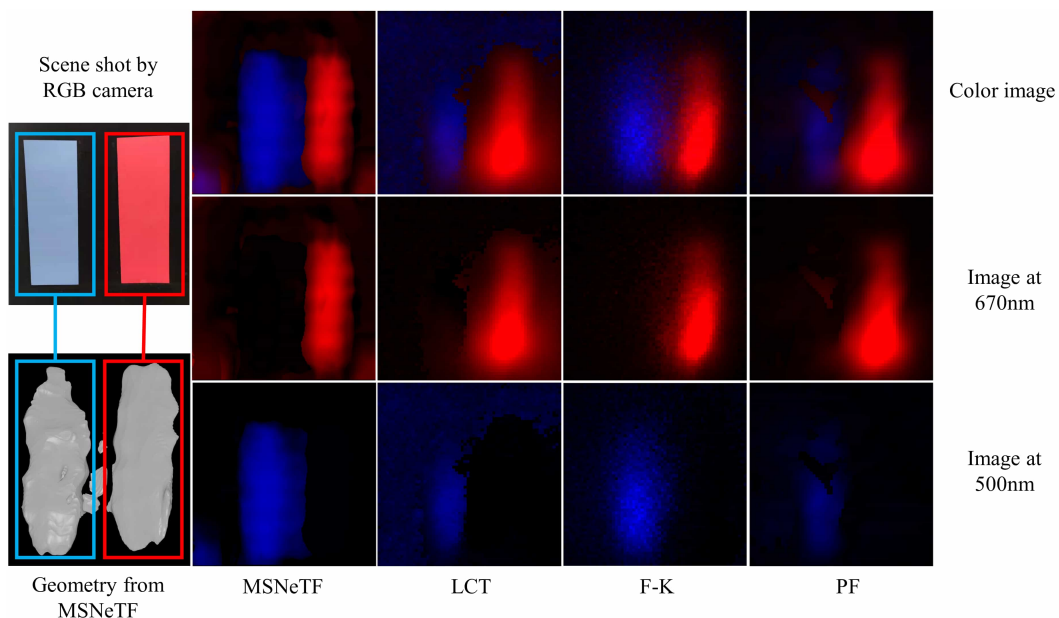


图 4-9 双色纸板上的结果
Figure 4-9 Results on red and blue bars.

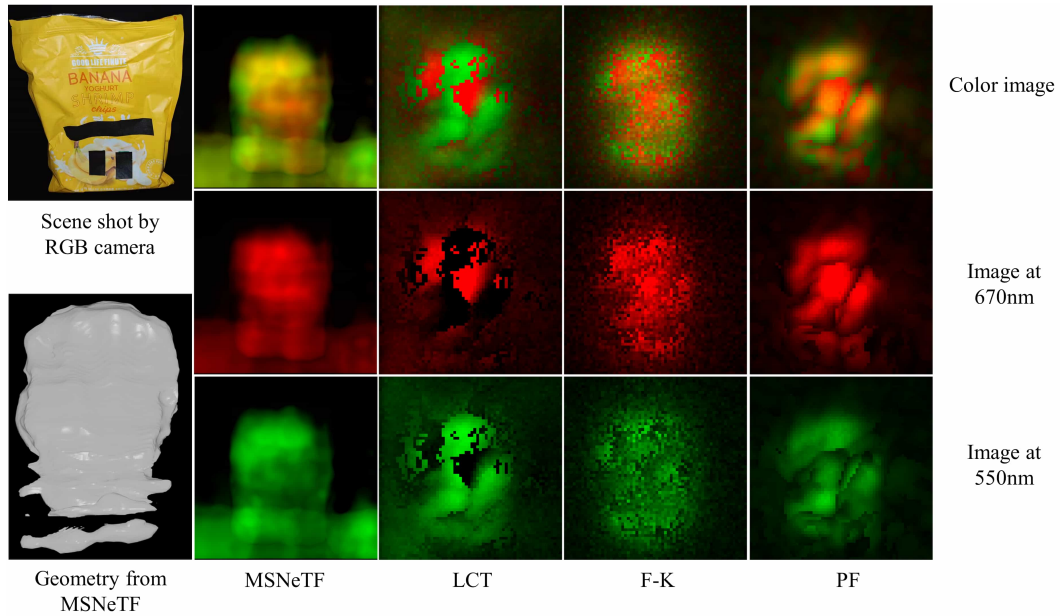


图 4-10 零食包装袋上的结果

Figure 4-10 Results on Snack.

性能。实验中，分别在 670 nm 和 550 nm 波长下采集瞬态数据，获得了对应红色（中行）和绿色（下行）的单通道图像，随后将二者融合生成彩色图像（上行）。

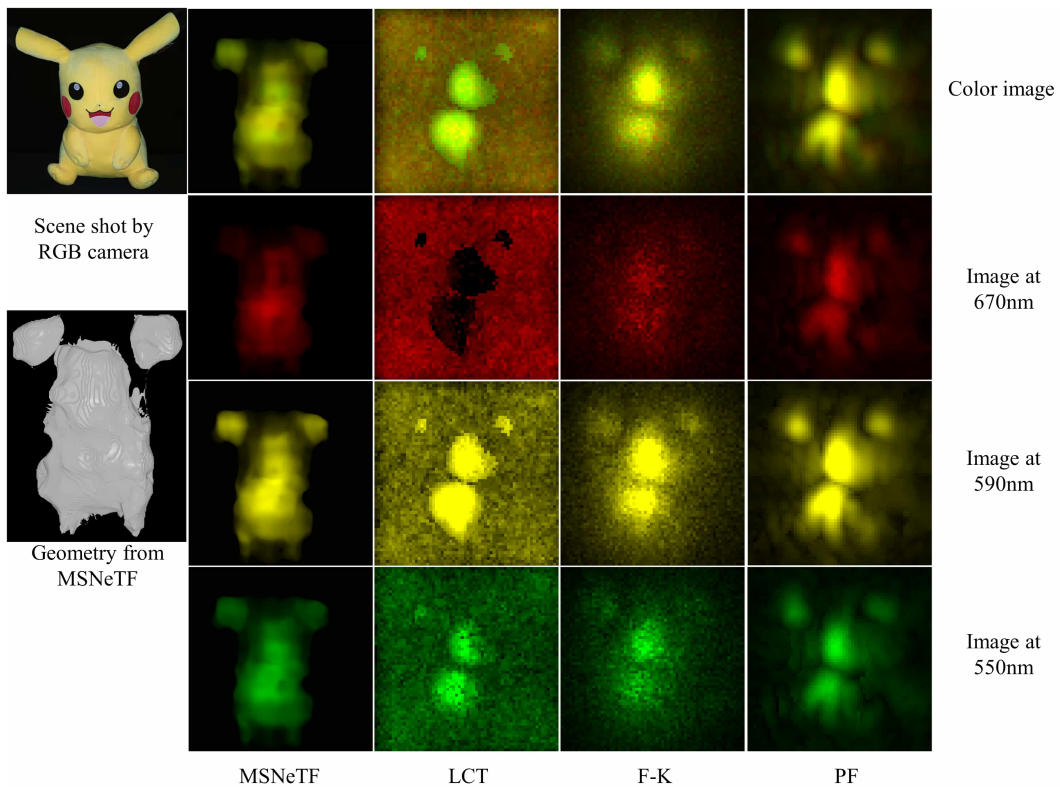


图 4-11 皮卡丘上的结果

Figure 4-11 Results on Pikachu.

如图 4-11展示了真实皮卡丘瞬态的重建，并对比了 SOTA 方法与 MSNeTF

的表现。实验中，在 670 nm、590 nm 和 500 nm 波长下分别捕获瞬态数据，对应生成红色（中间行）、黄色（下行）和绿色（底部）的单通道图像，随后将三通道数据融合生成彩色图像（上行）。该设计旨在充分展示 MSNeTF 在处理特定波长瞬态数据时的优势，同时深入探讨 SOTA 方法的性能表现。通过这样的比较分析，期望明确两种方法在重建过程中各自的优势与局限，为未来优化改进提供理论支持，并为应对类似挑战提供更为有效的策略。

4.5.5 定量比较

在本研究中，将表 4-1、4-2和4-3中展示的 MSNeTF 方法定量评估结果与 SOTA 方法进行了详细的对比分析。评估指标选用了学术界广泛认可的峰值信噪比、结构相似性和归一化均方根误差，以从不同角度反映图像质量。此外，还引入了深度平均绝对误差来量化几何重建误差，从而更准确地体现方法在形状恢复方面的性能。实验结果表明，MSNeTF 在图像质量和几何恢复上均优于 SOTA 方法，验证了其在处理此类问题时的优势。

表 4-1 贴片场景定量重建结果。

Table 4-1 Quantitative evaluation on patches.

	Confocal				Non-confocal			
	PSNR	SSIM	NRMSE	MAE*	PSNR	SSIM	NRMSE	MAE*
LCT	10.56	0.250	2.44e-6	0.049	10.61	0.235	2.31e-6	0.068
F-K	12.59	0.369	1.53e-6	0.046	10.19	0.222	2.66e-6	0.071
PF	9.05	0.191	3.46e-6	0.046	10.93	0.242	2.24e-6	0.068
MSNeTF	16.99	0.715	5.56e-7	0.046	14.16	0.443	1.07e-6	0.068

注：红色：最佳表现。

表 4-2 兔子场景的定量重建结果。

Table 4-2 Results on bunny.

	Confocal				Non-confocal			
	PSNR	SSIM	NRMSE	MAE*	PSNR	SSIM	NRMSE	MAE*
LCT	16.46	0.200	2.15e-6	0.0424	14.18	0.121	3.64e-6	0.121
F-K	14.57	0.165	3.33e-6	0.0379	14.24	0.116	3.59e-6	0.111
PF	15.68	0.272	2.58e-6	0.0293	13.50	0.107	4.26e-6	0.082
MSNeTF	17.73	0.759	1.61e-6	0.0150	16.75	0.718	2.02e-6	0.069

注：红色：最佳表现。

在本研究中，针对真实非视距场景提供了额外的重建结果，以丰富实验数

表 4-3 人体场景的定量重建结果。
Table 4-3 Results on human.

	Confocal				Non-confocal			
	PSNR	SSIM	NRMSE	MAE*	PSNR	SSIM	NRMSE	MAE*
LCT	15.15	0.027	4.01e-6	0.0538	13.07	0.014	6.48e-6	0.116
F-K	14.83	0.021	4.32e-6	0.0422	13.32	0.032	6.12e-6	0.140
PF	14.89	0.097	4.26e-6	0.0331	12.79	0.020	6.91e-6	0.232
MSNeTF	16.44	0.496	2.98e-6	0.0290	16.07	0.039	3.25e-6	0.037

注: 红色: 最佳表现.

据。实验选取了四类场景：由红蓝两平面条构成的 RBbar、穿着绿色内衣与红色外套的人体模型、黄色零食包装以及黄色玩具 Pikachu。这些场景在几何结构上由简至繁，且颜色覆盖蓝、绿、黄、红多个波长，确保了实验数据的多样性。

采用 MSNeTF 方法对这些非视距场景进行了重建，并与 SOTA 方法进行了详细对比。结果表明，MSNeTF 在恢复几何形状和反照率方面均优于 SOTA 方法，尤其在 Snack 和 Mannequin 等复杂场景中优势更加明显。此外，RBbar 和 Pikachu 场景的重建结果也支持 MSNeTF 方法在不同场景中的稳定性与鲁棒性。

总体而言，本研究的实验结果验证了 MSNeTF 在多种非视距场景重建中的优越性能，为进一步解决相关问题提供了有力的数据支持和理论依据。

表 4-4 不同数据的设定与参数
Table 4-4 Setup and parameter of different data.

Experiments	Mannequin	RBbar	Snack	Pikachu
NLOS size	0.6m×0.4m	0.45m×0.45m	0.32m×0.43m	0.5m×0.5m
NLOS distance	0.9m	0.9m	0.9m	1.0m
Detection region	0.7m×0.7m	0.7m×0.7m	0.7m×0.7m	0.7m×0.7m
Detection points	64 × 64	64 × 64	64 × 64	64 × 64
Wavelength ranges	530nm-570nm	480nm-520nm	545nm-555nm	545nm-555nm
	650nm-690nm	650nm-690nm	665nm-675nm	585nm-595nm
				665nm-675nm
Exposure time	1s	1s	1s	1s

4.6 结论

本文提出了一种基于多光谱神经瞬态场的非视距彩色成像技术。构建了多光谱非视距成像流程，并采用类似 NeRF 的深度学习框架实现跨光谱几何不变性。同时，深入研究 SPAD 噪声模型表明，使用泊松噪声有助于提升重建性能。

目前主要在合成数据上进行了验证，下一步计划在实际系统中的宽带激光条件下改进实验。未来的工作将探讨光谱间瞬态数据的相关性，例如通过稀疏采样减少测量，同时考虑在每个样本点获取不同光谱瞬态数据，并定制深度网络以解决几何重建和反射率估计问题。

第5章 基于预训练和光谱相位场的自监督学习光谱重建

5.1 引言

计算光谱成像在许多领域具有广阔的应用，但在实际应用中面临着很大的困难和挑战。一方面，由于工作环境的复杂性，光谱成像的前向模型在许多场景下是非局部的，而目前缺少对全局光谱成像模型的处理手段。并且，当前的光谱成像技术多基于硬件设备，这些设备自身的限制，例如分辨率低、编码设计受限等问题，使得光谱分析的精度和效率无法大幅度提高。此外，尽管神经网络的引入为提高光谱分辨率开辟了新的途径，但由于训练数据不足，神经网络在光谱成像中的应用同样受到限制。

针对这些问题，本文提出了一种创新的基于预训练和光谱相位场的自监督学习光谱重建方法。这种方法通过先验知识引导神经网络的预训练，使得光谱解码器能够对光谱信息进行精确解析。同时，本文引入了光谱相位场的概念，并将其纳入神经网络的训练中。通过这种方式，本文的方法不仅能够适应光谱数据的稀疏性，还能够利用预训练的解码器优化光谱相位场，降低噪声的干扰，进一步提高光谱成像的质量。

本文选择了具有挑战性的非视距光谱成像环境来验证本文提出的方法的有效性。非视距光谱成像技术为遮蔽环境下的物质识别与功能分析提供了革命性潜力，但其在数据采集时间和信号保真度方面仍面临严峻挑战。传统非视距光谱成像方法依赖于可调谐激光器的逐波长扫描机制，存在采集时间长、特定波段信噪比低等固有缺陷。现有计算方法通常独立处理单色数据，进一步加剧了噪声干扰和系统效率低下的问题。借助于本文提出的基于预训练和光谱相位场的自监督学习光谱重建方法，将多光谱滤光片阵列（Multi-Spectral Filter Array, MSFA）扫描策略与神经光谱相位场（Neural Spectral Phasor Fields, NeSPF）算法有机结合。MSFA 设计实现了多波长同步采集，而 NeSPF 算法则集成了隐式神经表示以实现光谱-空间联合重建，通过预训练光谱解码器引入先验知识，并采用迭代相位场优化有效抑制噪声。基于共焦单光子激光雷达系统的实验验证表明，与传统方法相比，本框架在 400-700 nm 波段（10 nm 分辨率）的采集时间缩短了 30 倍。相比现有技术，该方法在重建质量方面表现出显著优势，并通过真实与合成样本测试验证了其卓越的物质鉴别能力。这一突破性进展解决了非视距光谱成像长期存在的速度、信噪比和光谱保真度等问题，为隐蔽场景的非侵入式分析提供了实用化解决方案。

5.1.1 非视距光谱成像

非视距（Non-line-of-sight, NLOS）成像技术使得可以观察到被直接视线阻挡的物体。通过运用如光的飞行时间（Time-of-Flight, TOF）、散射相关性^[122]和波前塑造^[123,124]等非传统原理，人们已经取得了显著的进步。特别地，基于

TOF 的方法在多种场景下已经展示出了强大的三维重建能力,使其成为最有前景的解决方案之一^[125-137]。

基于 TOF 的方法通常使用单色超快脉冲激光束,将其定向至直接视野内的反射墙。该墙将激光束散射成一个向隐藏物体传播的球形波前。当波前与场景交互并反射回到墙上时,一个具有飞秒级分辨率的飞行时间探测器,如单光子雪崩二极管 (Single Photon Avalanche Diode, SPAD)^[117,138],会记录返回光子的到达时间。在时间相关单光子计数 (Time-Correlated Single Photon Counting, TCSPC) 模式下操作的 SPAD 传感器会生成一个瞬态直方图,其中反射墙上的单个点对应于激光波长处光子计数与时间间隔的直方图。

然而,由于存在一些挑战,对于进阶材料识别或功能成像至关重要的光谱 NLOS 成像仍然没有得到广泛的研究。首先,它需要一个可调的激光源,并在不同的波长上进行重复扫描,这大大延长了数据获取时间。其次,尽管单色 NLOS 成像通常使用的是高功率的激光源,但可调激光源在特定波长上的功率密度通常要低得多,导致 NLOS 光谱数据中的信噪比极低。

计算算法对于从光谱 NLOS 数据中重建光谱图像至关重要^[109,116,139-149]。一种方法是将光谱 NLOS 问题视为独立的单色 NLOS 成像问题,对每个波长应用适当的单色 NLOS 重建算法。过去的研究已经成功地使用单色飞行时间系统重建了隐藏物体的形状和反照率。开创性的滤波反投影 (Filtered Back-Projection, FBP) 及其变体将 NLOS 前向过程模拟为线性算子,并在测量数据上使用伴随算子重建隐藏的对象^[150-152]。另一种基于光锥变换 (Light Cone Transform, LCT) 的方法将 NLOS 前向模型简化为 3D 卷积,其中,像 Wiener 过滤器这样的高级信号处理技术可以进一步降低噪声^[119,153]。相位场 (Phasor Fields, PF) 方法认为 NLOS 成像问题可以通过衍射理论方法解决,并且引入带限虚拟相位场可以有效抑制 NLOS 数据中的宽带噪声^[130,154,155]。信号-对象协同正则化 (Signal-object collaborative regularization, SOCR) 方法考虑了对重建目标和测量信号的先验知识,即使在粗糙和不规则的 NLOS 数据下也能产生高质量的重建^[118,156]。NeTF 方法将神经网络隐式表达引入 NLOS 问题,提供了内存效率与诸如深度先验等其他模块的简单集成^[157,158]。尽管最近取得了一些进展,但直接在每个期望的波长上应用这些方法进行 NLOS 光谱重建会显著增加获取时间,并导致非常低的 SNR,从而导致成像性能不佳。因此非视距光谱成像仍然是一个待解决的问题。

5.1.2 非视距光谱成像系统

图 5-1 展示了三种不同的光谱非视距成像框架的流程。**a** 展示了传统的非视距成像使用单色光探测隐藏物体,收集单色时间直方图。然后应用重建算法来恢复隐藏场景的灰度表示。**b** 展示了全光谱非视距成像使用多波长的光依次照射隐藏物体,获取一系列特定波长的时间直方图。重建算法独立应用于每个直方图,生成每波长的图像,然后将它们组合成一个光谱图像栈。**c** 展示了所提出的快照

光谱非视距成像系统利用设计的照明模式探测隐藏物体，捕获与传统单色采集相当的数据量的稀疏光谱直方图。通过利用神经光谱相位场，该框架重建了与全光谱方法相当的的质量的光谱图像，同时保持了与单色非视距成像相当的数据采集时间。

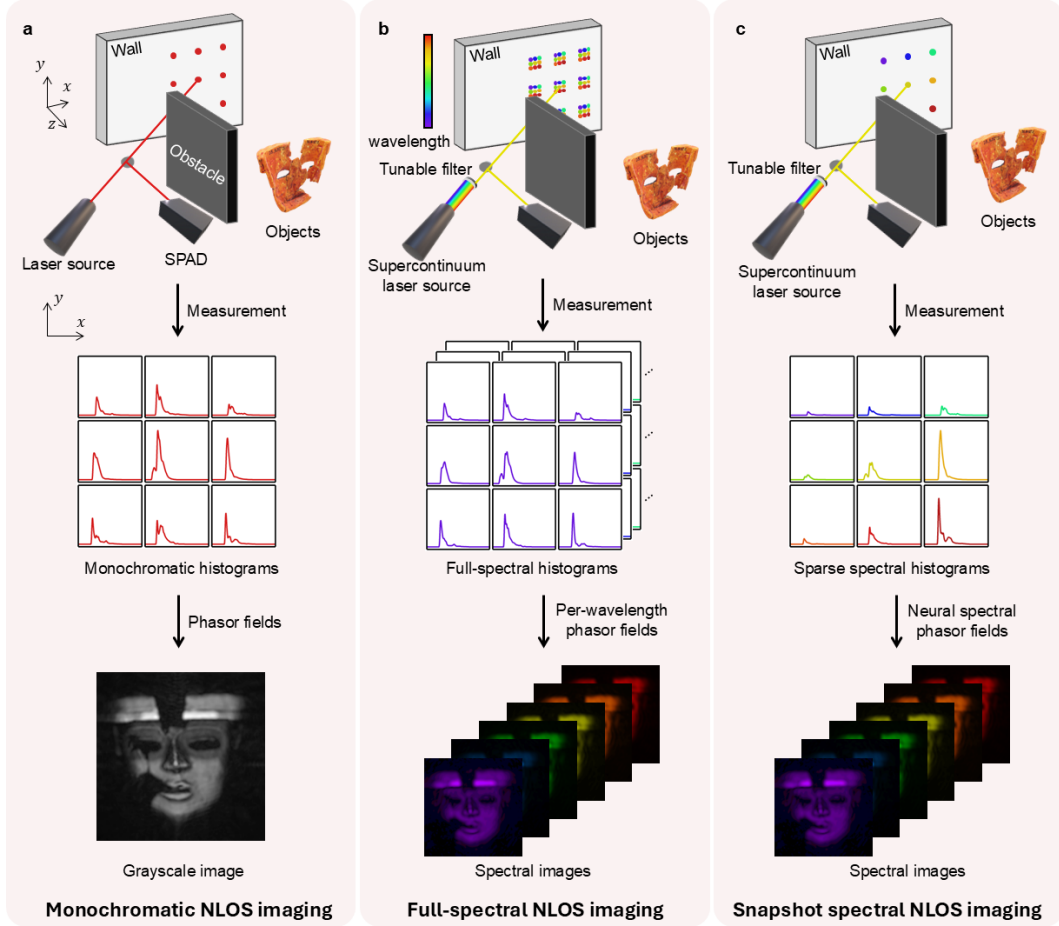


图 5-1 三种不同的光谱非视距成像框架

Figure 5-1 Pipeline of three different spectral NLOS imaging framework.

5.2 前向模型

快照式非视距光谱成像通过采集稀疏的光谱光子计数，通过最小化测量光子计数和非视距光谱前向模型预测的光子计数之间的差异，重构隐藏对象的光谱反射率。

首先分析并离散化非视距光谱正向模型。设 $\mathbf{x} = (x, y, z)$ 表示隐藏对象上一点的坐标， $\rho(\mathbf{x}, \lambda)$ 表示其在波长 λ 处的光谱反射率。正向模型可表达为：

$$\tau(\mathbf{x}_p, \lambda, t) = \int_{\Lambda} \int_V \frac{\mathcal{L}(\mathbf{x}_p, \lambda) S(\lambda)}{|\mathbf{x} - \mathbf{x}_p|^4} \rho(\mathbf{x}, \lambda) \delta\left(|\mathbf{x} - \mathbf{x}_p| - \frac{ct}{2}\right) d\mathbf{x} d\lambda, \quad (5-1)$$

其中, $\mathbf{x}_p = (x_p, y_p, 0)$ 对应于中继墙面上的一个扫描点, $\tau(\mathbf{x}_p, \lambda, t)$ 是在波长 λ 处由 SPAD 记录的光子计数-时间直方图, Λ 表示光谱范围, V 定义了物体的潜在空间域, $\mathcal{L}(\mathbf{x}_p, \lambda)$ 是在 $\mathbf{x}_p$ 处的激光强度, $S(\lambda)$ 表示 SPAD 的光谱响应, δ 是狄拉克 δ 函数。

光子计数直方图和反射率之间的线性关系, 因此可以将全光谱成像模型离散化为:

$$\boldsymbol{\tau} = A\boldsymbol{\rho}. \quad (5-2)$$

这里, $\boldsymbol{\tau} \in \mathbb{R}^{MCT}$ 包括离散化的光子计数直方图, 其中 M 是单色扫描时扫描点的数量, C 是光谱通道数, T 是光子计数直方图的长度。因此, 全光谱成像的总扫描点数等于 MC 。通过将对象区域划分为 N 个体素, 得到离散化的反射率 $\boldsymbol{\rho} \in \mathbb{R}^{NC}$ 。 $A \in \mathbb{R}^{MCT \times NC}$ 表示离散化的前向模型。

为了避免全光谱扫描, 提出在特定的波长扫描模式下照明每个中继墙面点, 从而实现稀疏的空间-光谱采样。这产生了快照式的前向模型:

$$\boldsymbol{\tau}_s = \omega_s A \boldsymbol{\rho}, \quad (5-3)$$

其中, $\omega_s \in \mathbb{R}^{MT \times MCT}$ 中只有 0 列和非 0 列, 意味着从全光谱成像模型中选择部分行。通过扫描 M 个点而不是 MC 个点, 采集时间减少了 C 倍, 这与单色 NLOS 成像是一样的。

然而, 从稀疏的光谱数据中重构 4D (3D 空间 + 1D 光谱) 反射率构成一个严重病态的反问题。此外, 如何设计波长扫描的模式以实现高效的数据采集和重建仍是一个待解决的挑战。

5.2.1 扫描模式设计

本文中的扫描模式设计受到数字相机中光谱滤波器的启发^[159]。扫描模式设计遵循两个关键要求: 1) 空间均匀性: 为解决每个空间点的单波段采样, 必须从相邻的采样点推断出未测量的光谱组分。这就需要在中继墙上每个光谱带的空间分布近乎均匀。区域之间的像素密度偏斜可能导致在插值过程中丢失大量信息, 从而损害重建的完整性。2) 通道均匀性: 虽然先前的研究优先将额外的像素分配给目标特定的光谱带以增强识别, 但提出的方法强调通过均匀化的通道采样概率来提高光谱的可靠性。每个光谱带都保持平衡的空间表示, 以避免在光谱估计过程中产生偏差。

实施了一个迭代优化框架, 包括两个顺序操作: 1) 分解步骤: 将棋盘格基本模式分解为互补的二进制层 (0/1 掩码), 每一层都对应于不同的光谱带。2) 子采样步骤: 对每个二进制模式应用密度感知的下采样, 保持均匀的空间覆盖, 同时减少层特定的采样密度。

Fig. 5-2说明了构造支持 8 个光谱通道的 4×4 网格的过程。初始的多尺度分解通过迭代的棋盘格分割生成二进制子模式，每次迭代应用位逻辑左移操作以编码光谱索引。子采样对这些模式进行精细化处理，以确保每个通道的空间均匀性。从连续的分解层中聚合的二进制编码被转换为十进制值，定义每个扫描点的光谱通道分配。这种分层的方法保证了成像网格上的空间规律性和通道平等性。

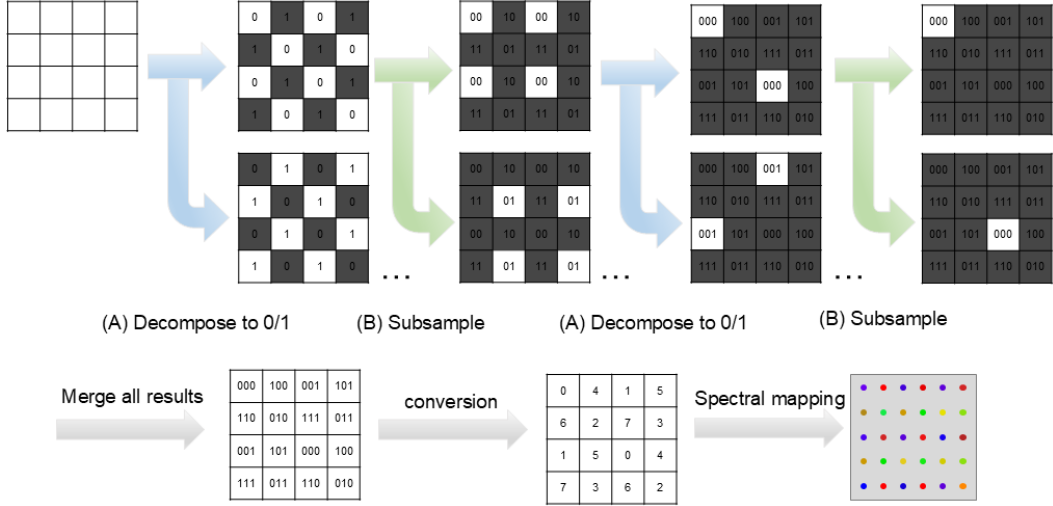


图 5-2 扫描模式设计

Figure 5-2 The design process of MSFA scanning pattern.

5.3 神经光谱相位场

本文提出了神经光谱相位场，一个从稀疏光谱瞬变中重构光谱图像的框架。NeSPF 框架整合了三个核心组件：(1) 相位场调节的神经场以增强噪声鲁棒性，(2) 预训练的光谱解码器强制空间-光谱先验，和 (3) 在光谱特征而非反射率上操作的正向模型以增加灵活性。如图. 5-3所示，提出的方法首先通过将相位场应用于稀疏光谱直方图，来估计初始光谱图像。这些初始估计的空间坐标和对应的强度值作为输入进入一个多层感知机，作为一个条件神经场。这个多层感知机与一个预训练的光谱解码器串联，生成光谱特征，这些特征被集成到一个前向模型中，预测光谱直方图以优化网络参数。在训练收敛后，预测的全光谱直方图通过相位场进行处理，重建隐藏物体的高保真度光谱图像。

NeSPF 流程 (图 5-3) 从初始的光谱图像估计 $I_0(\mathbf{x}, \lambda)$ 开始，从测量的快照光谱数据 $\tau_m \in \mathbb{R}^{M \times T}$ 中插值并通过相位场增强生成：

$$I_0(\mathbf{x}, \lambda) = PF \left(Interp(\tau_m) \right). \quad (5-4)$$

在这里， $\tau_m \in \mathbb{R}^{M \times T}$ 表示测量的快照光谱数据。对于每个空间坐标 $\mathbf{x} = (x, y, z)$ 和波长 λ ， I_0 代表隐藏物体的光谱反射率的初始估计。随后，利用条件

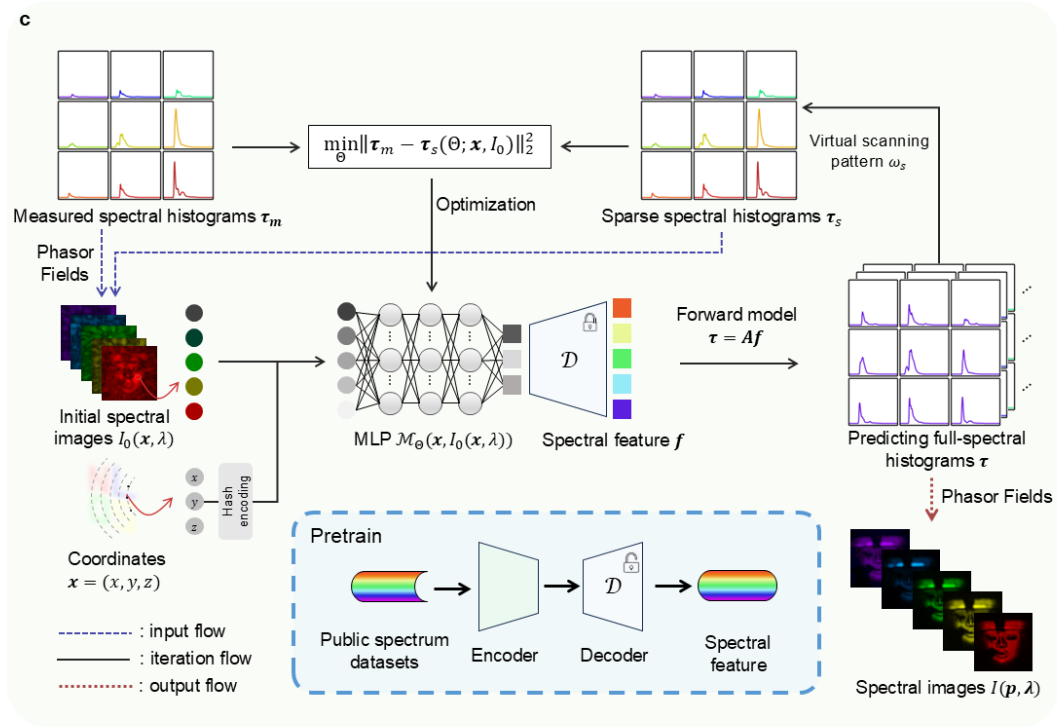


图 5-3 神经光谱相位场的方法流程

Figure 5-3 Pipeline of neural spectral phasor fields.

神经场，它条件于隐藏物体的估计光谱反射率。估计的光谱反射率及其对应的通过哈希编码编码的坐标作为输入传入多层感知器 $\mathcal{M}_\Theta$ ，其中 Θ 表示 MLP 的参数。MLP 在坐标 $\mathbf{x}$ 处输出一个潜在的表示。

随后，坐标 $\mathbf{x} = (x, y, z)$ 和他们的哈希编码光谱强度 $I_0(\mathbf{x}, \lambda)$ 被输入到多层感知器 $\mathcal{M}_\Theta$ ，产生潜在表示。这些表示被解码成光谱特征 $f(\mathbf{x})$ ，使用预训练的光谱解码器 $\mathcal{D}$ ，在公共光谱数据集上训练，以包含自然光谱的空间-光谱相关性：

$$f(\mathbf{x}) = \mathcal{D}(\mathcal{M}_\Theta(\mathbf{x}, I_0(\mathbf{x}, \lambda))). \quad (5-5)$$

使用光谱特征替代成像模型中的反射率 (Eq. 5-1)，以进行光谱数据的预测：

$$\tau = Af. \quad (5-6)$$

通过最小化预测和测量的光谱数据之间的 ℓ_2 -范数差异来优化 MLP 参数 Θ ：

$$\Theta^* = \arg \min_{\Theta} \|\tau_m - \omega_s A \mathcal{D}(\mathcal{M}_\Theta(\mathbf{x}, I_0(\mathbf{x}, \lambda)))\|_2^2. \quad (5-7)$$

在优化过程中，神经光谱相位场使用更新的光谱数据预测来迭代优化地输出光谱图像。最终在优化完成后，通过将相位场应用到预测的全光谱数据，得到最终的重构光谱图像。

5.3.1 瑞利-索末菲衍射相位场

相位场是一种快速且鲁棒的非视距成像工具。在流程中，相位场在开始时用于从稀疏光谱数据中提取初始的光谱辐射场，在最后用于从重构的全光谱数据中恢复精确的光谱辐射场。

相位场的工作原理是将光谱非视距数据 $\tau(\mathbf{x}_p, \lambda, t)$ 物理上视为从中继墙面上的点 $\mathbf{x}_p$ 发出的单色脉冲信号的隐藏物体的脉冲响应。

$$\mathcal{P}_r(\mathbf{x}_p, \lambda, t) = \int \mathcal{P}_e(\mathbf{x}_p, \lambda, t) * \tau(\mathbf{x}_p, \lambda, t) d\mathbf{x}_p \quad (5-8)$$

其中， $\mathcal{P}_r(\mathbf{p}, \lambda, t)$ 是在 $\mathbf{p}$ 处的相位场。 $\mathcal{P}_e(\mathbf{p}, \lambda, t)$ 是虚拟的复值单色发射信号，可以定义为：

$$\mathcal{P}_e(\mathbf{p}, \lambda, t) = e^{i\Omega_C t} \delta(\mathbf{p} - \mathbf{x}_0) \delta(\lambda) e^{-\frac{t^2}{2\sigma^2}} \quad (5-9)$$

虚拟的中心频率 Ω_C 根据空间中中继墙面上点的采样频率选择。时间延迟 $\delta(\mathbf{p} - \mathbf{x}_0)$ 表示光从光源到中继墙面的距离。并且应用 $e^{-\frac{t^2}{2\sigma^2}}$ 来模拟带宽限制的脉冲。

理论表明，可以通过与照明波 $\mathcal{P}_e(\mathbf{p}, \lambda, t)$ 卷积，就好像中继墙面后存在一个虚拟的直接成像的相机，因此可以应用 RSD 的波传播理论进行非视距成像。

$$I(\mathbf{x}, \lambda) = \Phi(\mathcal{P}_e(\mathbf{p}, \lambda, t) * \tau(\mathbf{p}, \lambda, t)) \quad (5-10)$$

其中， $R(\mathbf{x}, \lambda)$ 是在 $\mathbf{x}$ 和波长 λ 处的非视距图像。PF-RSD 对非视距成像过程中的噪声具有鲁棒性，提出的方法中融合了这种算法，以处理光谱非视距成像中的噪声。

5.3.2 神经场中的光谱解码器

提出的方法的核心部分是一个带有预训练光谱解码器 $\mathcal{D}$ 的基于坐标的 MLP $\mathcal{M}_\Theta$ 。MLP 将隐藏对象上的 3D 坐标 $\mathbf{x}$ 映射到相应的潜在表示，然后由预训练的解码器解码成光谱特征 f 。

在将它们输入到 $\mathcal{M}_\Theta$ 之前，将坐标网格归一化到立方体 $[-1, 1]^3$ 。神经网络的输入坐标首先被哈希编码，这对于 MLP 表示高频变化和神经网络在训练期间的快速收敛已被证明是必要的。网络 $\mathcal{M}_\Theta$ 由 4 个全连接层组成。前 3 层有 64 个由 leaky ReLU 激活的隐藏神经元，而最后一层有 64 个未激活的隐藏神经元。在第 3 个全连接层实现了跳跃连接，将 $\mathcal{M}_\Theta$ 的原始输入与中间输出连接起来，这已经被证明有利于提高表示性能。然后， $\mathcal{M}_\Theta$ 的输出是潜在表示。

然后，从 $\mathcal{M}_\Theta$ 输出的潜在表示被输入到预训练的光谱解码器中，以纳入自然光谱的光谱先验。光谱解码器是一个 MLP，由 4 个全连接层组成，其中前 3 层由 ReLU 激活并通过批量归一化层进行归一化。最后一层没有装备激活函数和

归一化层。在纳入非视距光谱成像之前，光谱解码器在公开的光谱数据集上进行预训练。然后，光谱解码器被应用于 NLOS 光谱成像，并输出光谱特征 f 。

在公开的光谱数据集上对自编码器进行预训练，并使用预训练的解码器进行的非视距光谱成像。自编码器包含了一个编码器和一个解码器，每个都有 4 个隐藏层。每个隐藏层有 128 个神经元，并通过 relu 函数激活。选择了所有来自 ICVL (Imperial Computer Vision & Learning) 数据集的实验室场景数据，包含校准板进行自编码器的训练。使用场景中的校准白板的光谱对光谱图像进行校准。然后将每个像素的校准光谱添加到数据集中。使用 MSE 作为损失函数，adam 优化器来训练自编码器。在 10000 个 epoch 之后，自编码器训练完成。然后只使用解码器进行非视距光谱重构。

5.4 实验设置

在 Nvidia RTX 3090 GPU 上训练神经光谱相位场模型。最大训练步骤设定为 200k。在这个设置下，每个 NLOS 场景的模型训练大约需要 3 小时。在每一步，选择一批大小为 2 的直方图来优化网络。对于每个直方图，在对象空间中以 $32 \times 32 \times 512$ 点进行采样，并从 PF-RSD 插值光谱强度，然后将这些坐标和光谱强度输入到 MLP。实现了一个递减的学习率，它随着训练 epoch 的增加呈指数衰减，以平滑优化。初始学习率设定为 1×10^{-3} ，最小学习率设定为 1×10^{-5} 。

实验设置显示在图 5-4 中。在实验中使用的光源是一种超连续激光器 SuperK 299 EXTREME FIU-15，重复频率为 19 MHz。为了生成可调窄带激光，应用了一个 SuperK VARIA 可调波长滤波器。在实验中使用的带宽为 10 nm，从 450 nm 变化到 750 nm，间隔 10 nm，总共有 31 个光谱带。激光和滤波器的输出平均功率为 1 mW，带宽为 10 nm，重复频率为 19 MHz。激光和滤波器的总时间抖动约为 180 ps。一个 2 轴扫描振镜 (Thorlabs GVS212) 在距系统约 1m 的墙上对照明和检测点进行光栅扫描。检测到的光由 50mm 佳能 f/1.8 镜头聚焦。时间分辨探测器是 Micro Photon Devices 的 FastGatedSPAD，具有 $50 \mu\text{m} \times 50 \mu\text{m}$ 的有效面积，报告的时间抖动为 50 ps，光子检测效率在 400nm 时为 50%，在 600nm 时为 20%，在 800nm 时为 5%。检测事件使用 PicoHarp 300 TCSPC 模块在 4ps 的时间分辨率下进行时间戳。硬件设备位于距中继墙面 1.5m 的位置，NLOS 场景隐藏在直接视线之外。视场为 40° 。墙是电影幕布，具有良好的漫反射性能和平滑的光谱反射率。SPAD 聚焦在网格中心附近的位置。避免扫描 SPAD 聚焦位置（共焦位置）周围的小方区域，因为来自墙的直接反射的信号强而无法被门控，导致直方图含有高的噪声。

系统在中继墙面上扫描一个 128×128 的网格，每个点对应于由 MSFA 模式定义的波长。由于振镜的转换速度远比光谱滤波器的响应速度快，依次扫描所有具有相同波长的点（例如，首先是 400 nm，然后是 410 nm，等等）。在 19.5 MHz 的重复频率下，PicoHarp 300 记录了由 12,808 个时间区域组成的原始直方图，时间分辨率为 4 ps。为了隔离信号光子，采集到的含有隐藏物体的直方图减

去了没有物体的参考测量，消除了暗计数和激光功率泄漏的贡献。为了将光子计数在时间轴上对齐，将振镜的直接反射的脉冲定位在时间 $t = 0$ ，然后截断到 4,096 个时间区域。每个光子计数直方图需要 1 秒的采集时间。

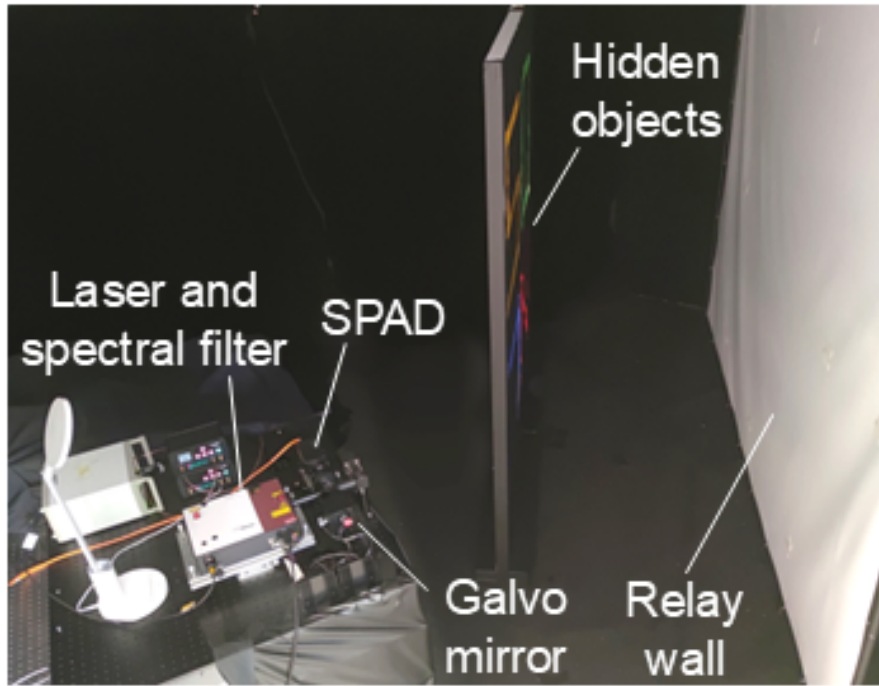


图 5-4 非视距光谱成像系统

Figure 5-4 Experimental setup.

5.5 成像系统

5.5.1 光谱和空间定标

光谱定标包括校准不同波长的光谱强度和时间延迟。通过测量具有已知光谱反射率的白板，校准了光谱非视距成像系统（包括激光器、可调滤波器、分束器、振镜、中继墙面、镜头和探测器）的总光谱响应。假设在扫中继墙面上的每个点时，系统的光谱响应和时间延迟保持不变，因此只需要测量一个扫描点的白板的光谱响应和时间延迟。为了在非视距设置下测量白板，选择这个扫描点位于白板在中继墙面上的投影的中心。在从 450nm 到 750nm 的波长上，间隔 10nm，测量了光谱非视距的光子技术直方图。然后通过时间维度上的高斯滤波器对光谱非视距直方图进行平滑，以减少泊松噪声。然后在时间维度上找到直方图的最大值，并获得光谱强度。也找到最大值出现的时间，并获得时间延迟。时间延迟需要减去延迟向量的最小值，使得延迟的最小值为 0。测得的光谱强度和时间延迟的尺寸都是 31×1 。光谱光子计数、光谱强度和时间延迟显示在图 5-5 中。在光谱定标之后，当为一个隐藏的物体测量一组光谱数据时，需要通过标定时测量的光谱强度向量对待测目标的直方图进行归一化，以获得隐藏物体的真实光谱反射率，并通过测量的延迟向量补偿不同波长的延迟。

硬件系统的空间定标需要对齐激光和探测器的点，并通过振镜的电压控制点的扫描。首先，调整 SPAD 探测器的位置和角度，以在低激光功率和 SPAD 在自由运行模式下最大化光子计数。然后，对系统进行误差对准，以便在墙上照射和成像两个略有不同的点，以减少在 NLOS 测量中来自中继墙面的直接光。其次，在中继墙面上栅格扫描 13×13 个点，振镜的电压是均匀的。然后可以估计振镜的位置和中继墙面的方程。还使用这些信息在扫描过程中调整 SPAD 门控开启的时间，使得能够拒绝测量过程中来自在墙的直接反射。振镜的伏特/度缩放因子在扫描具有高宽度距离比（例如，1.6m 宽度和 1m 距离。）的场景时具有非线性。因此，额外扫描一个 5×5 棋盘校准目标，并在 NLOS 测量期间补偿非线性。

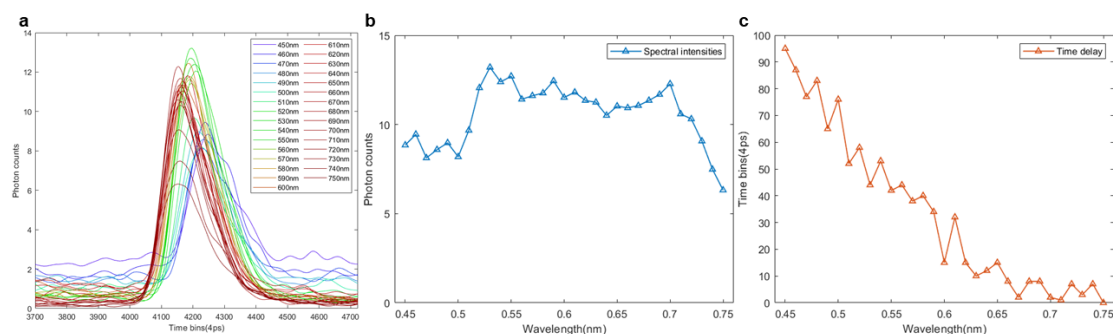


图 5-5 非视距光谱成像系统光谱定标

Figure 5-5 Spectral calibration of the imaging system using a reference white board with known spectral reflectance.

5.5.2 激光功率泄漏

使用一个超连续激光源和一个可调滤波器，以实现在各种波长和特定重复频率下的窄带激光束。然而，可调滤波器的不完善可能导致在不同的波长和重复频率下的激光功率泄漏，导致显著的后脉冲噪声。如图 5-6所示，在两种单光子雪崩二极管条件下进行了 NLOS 成像实验——自由运行模式和门控模式，分别在 450 nm 和 550 nm 的波长下。图 5-6(a) 展示了实验设置。两个振镜实现了二维扫描，导致在直方图开始时有两个振镜反射峰。调整直方图在时间上的偏移量，使得第一个振镜反射峰对准 $t = 0$ 。重复频率设定为 19.5 MHz，直方图的时间长度为 12808，每个时间区域对应 4 ps。除了中继墙面反射峰和非视距信号外，还有一组振镜峰，由于在 39 MHz 和 78 MHz 处的激光功率泄漏。这种泄漏是由于可调滤波器的不完美的阻带性能造成的。在 450 nm 处能够观察到泄漏的激光功率，但在 550 nm 处没有观察到，如图 5-6(b) 所示。图 5-6(c) 和 (d) 说明了当 SPAD 在门控模式下运行时的直方图。

功率泄漏影响了从光谱非视距成像实验中测量的直方图的两个重要方面。首先，高重频的振镜峰比非视距信号和噪声的光子计数高大约 100 倍，导致了大量的数据噪声。其次，振镜峰的后脉冲产生的高斯噪声大约是非视距信号的三倍。解决第一个问题，由于高重频的振镜峰相比于非视距信号和噪声有极高的光子

计数，所以这些峰可以被轻易地检测和定位。并且由于这些峰是来自于系统内部，与物体或中继墙面无关，有固定的时间位置和宽度，所以可以被识别并从基于优化的框架中排除出去。对于第二个问题，后脉冲类似于高斯噪声，这是各种单光子成像设置中常见的。因此，依赖于框架中集成的相量场来解决这个问题。

激光功率泄漏的问题源于激光和可调滤波器的不完善，而不是提出的方法的缺陷。相信随着激光技术的进步，这个问题可以有效地解决。

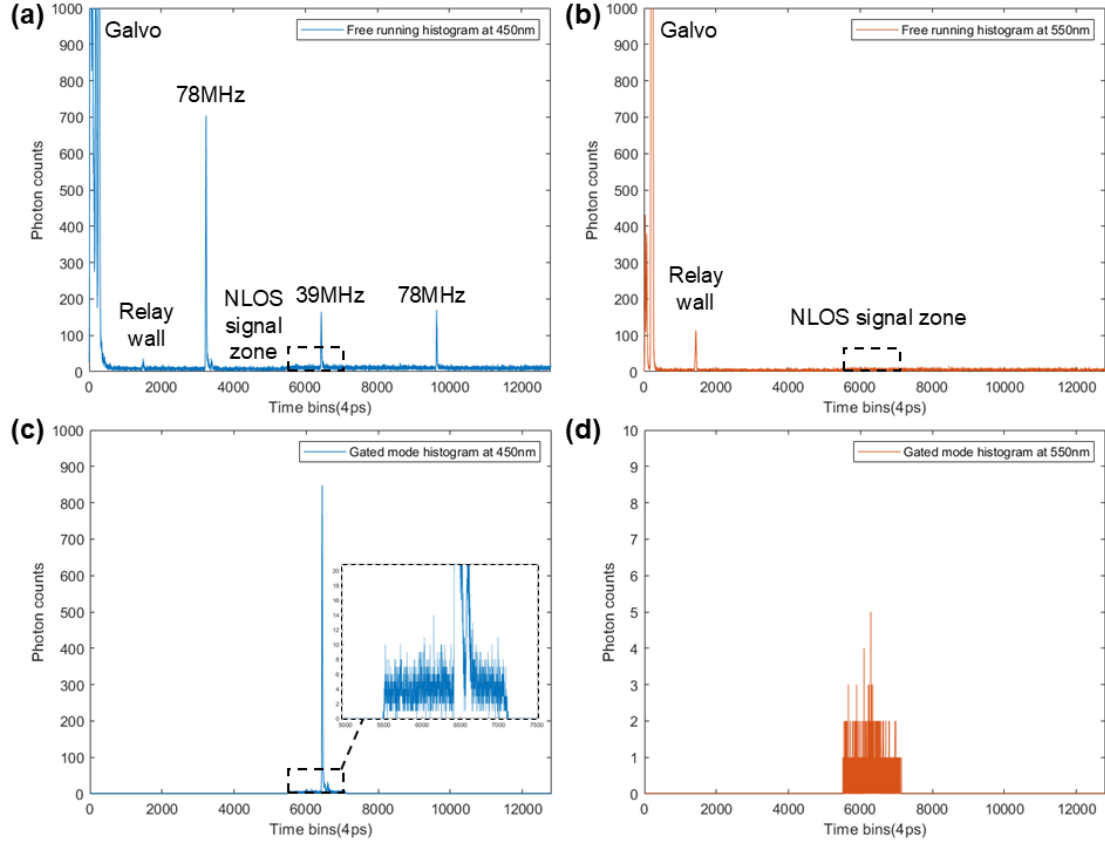


图 5-6 超连续激光系统中的功率泄漏

Figure 5-6 Illustration of laser power leakage in a 19.5 MHz laser system, demonstrating power leakage at 39 MHz and 78 MHz at 450 nm.

5.6 实验结果

5.6.1 仿真数据结果

首先使用一个包含四个窄带光谱条的合成场景来验证提出的框架。每个条均在 460 nm、530 nm、570 nm 和 650 nm 处表现出不同的反射峰，尺寸为 0.4 m (长度) $\times$ 0.04 m (宽度)，间距为 0.04 m。场景位于中继墙面 1 m 处，扫描范围为 2 m $\times$ 2 m (128 $\times$ 128 点)，时间分辨率为 32 ps，光谱分辨率为 10 nm (在 400-700 nm 范围内的 31 带)。

图. 5-7展示了对提出的方法和扫描模式在合成窄带光谱数据上的评估。**a** 展示了需要一个扫描模式来恢复隐藏的四条目标。**b** 展示了通过迭代子采样和分解

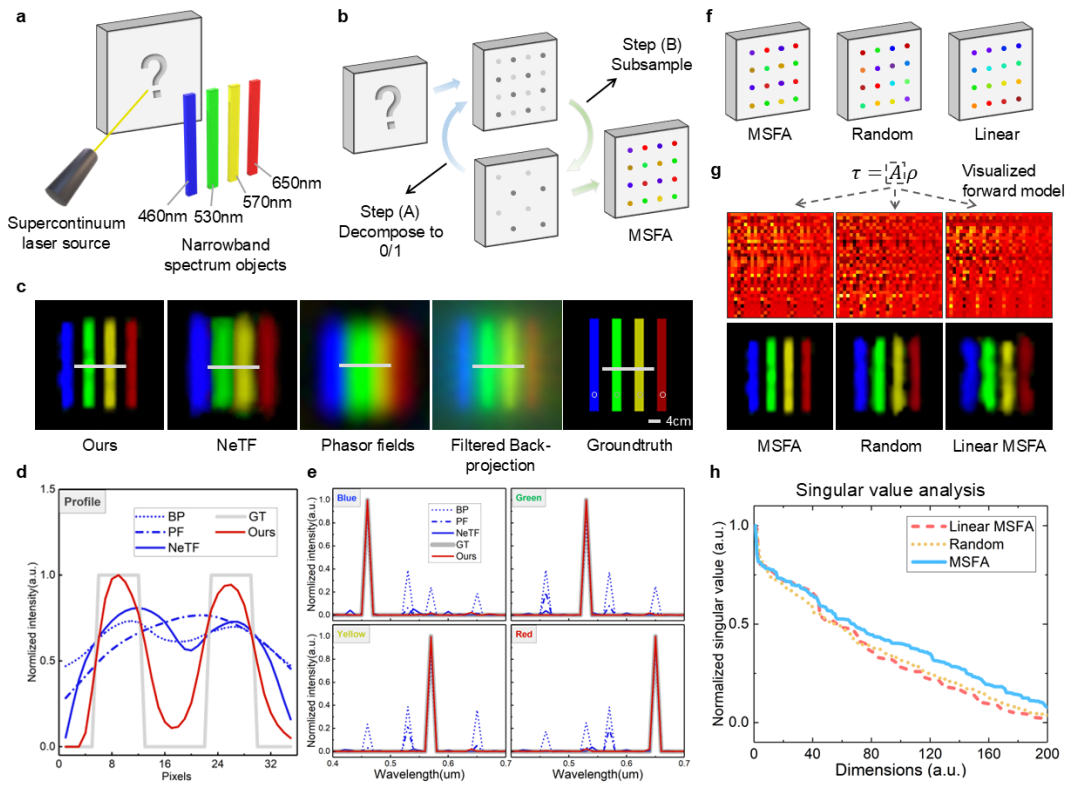


图 5-7 窄带反射率物体的重建

Figure 5-7 Evaluation of the proposed method and scanning patterns on synthetic narrow-band spectral data.

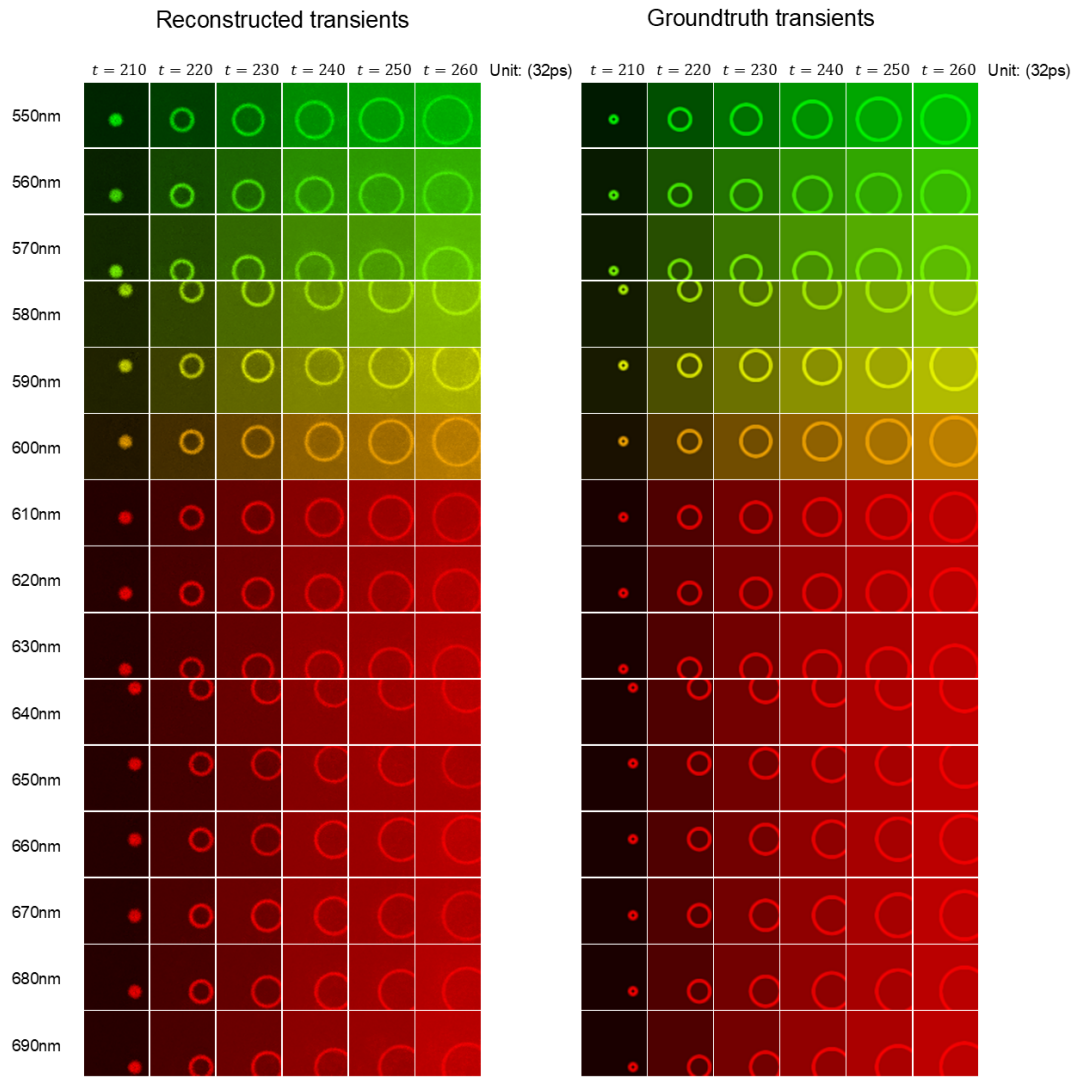


图 5-9 窄带反射率物体的完整光谱数据重建 (400nm-540nm)

Figure 5-9 Evaluation of spectral transients via our method on synthetic narrow-band patches. (400nm-540nm)

生成多光谱滤波器阵列。**c**展示了使用的方法、NeTF、相位场和反投影（合成为RGB 以便于可视化）的比较光谱图像重建。**d**展示了沿着所有方法选择的线恢复的空间轮廓。**e**展示了从每种方法中选择像素的恢复光谱。**f**展示了分析的三种扫描模式：MSFA、随机和线性配置。**g**展示了每种扫描模式的前向模型可视化和对应的重建。**h**展示了用于模式比较的前向模型奇异值分解分析。

为了设计扫描波长模式，采用了多光谱滤波器阵列策略，通过对初始二元模式的迭代子采样和分解，确保每个点均匀分配二元代码。与随机或线性模式相比，该方法可以减少成像偏差。

图 5-7将提出的方法与 NeTF、Phasor Fields (PF) 和 Filtered Back-Projection (FBP) 进行比较。NeTF 和 FBP 直接处理快照瞬态，而 PF 需要插值全光谱数据。在图 5-7c 中显示了归一化到 [0,1] 范围的合成 RGB 图像。重建结果揭示了 NeTF/PF 的模糊输出和 FBP 的低对比度伪影，而提出的方法能准确地定位目标。空间剖面图（图 5-7d）显示提出的方法与 PF 的未解析特征和 FBP/NeTF 的部分清晰度相比具有全分辨率。光谱重建（图 5-7e）显示提出的方法与地面真实的窄带峰非常接近，而其他方法展示了虚假的伪影。在 MSFA 模式下的定量分析得到了 PSNR 值：17.65（提出的方法），11.79（FBP），12.49（PF）和 13.36（NeTF）。随机（17.49）和线性 MSFA（15.22）模式的性能下降，强调了提出的设计的有效性。

进一步评估了不同的扫描模式（图 5-7f）：MSFA，随机（波长随机分配），和线性（波长均匀增加）。图 5-7g 展示了正向模型和重建结果，其中 MSFA 产生了与其他模式的降级输出相比的清晰、无伪影的图像。奇异值分析（图 5-7h）显示 MSFA 的 SVD 谱的衰减较慢，表明其相对于其他方法具有优越的条件。

进一步测试了一个合成的宽带场景，该场景包含一个由具有不同光谱的粘土类型组成的 $1.7\text{ m} \times 1.5\text{ m}$ 的古代面具（图 5-10）。提出的方法的重建（图 5-10a）与 PF 的全光谱参考匹配，而 NeTF 丢失了面部细节，FBP/PF 受到模糊或畸变的影响。在选定的波长下的光谱图像（图 5-10b）区分了材料性质，分解结果（图 5-10d）识别了三种粘土类型的空间分布。面具区域的全光谱重建（图 5-10c）与地面真实情况相符，得到了 PSNR 值：21.12（提出的方法），19.40（NeTF），17.70（FBP），和 17.54（PF）。

这些实验确认了提出的方法在空间分辨率、光谱保真度和对稀疏采样的鲁棒性方面的优越性，在窄带和宽带场景中均优于现有技术。

如图 5-11所示，使用由 30 个窄带反射贴片（各 $6\text{ cm} \times 6\text{ cm}$ ）组成的合成数据，评估提出的方法解决复杂场景的能力，每个贴片具有不同的光谱峰，范围从 400-690 nm。使用 MSFA 模式，扫描 $1\text{ m} \times 1\text{ m}$ 的场景区域。与 NeTF、Phasor Fields (PF)、无插值的 PF 和 Back-Projection (BP) 相比，重建的 RGB 图像和光谱剖面。提出的方法能够准确地定位所有的贴片，同时保持光谱保真度。NeTF 引入了空间扭曲，这些扭曲传播到光谱偏差，特别是在重叠区域。PF 保持贴片的几何形状，但由于插值伪影，颜色混合，而无插值的 PF 无法恢复形状和光谱

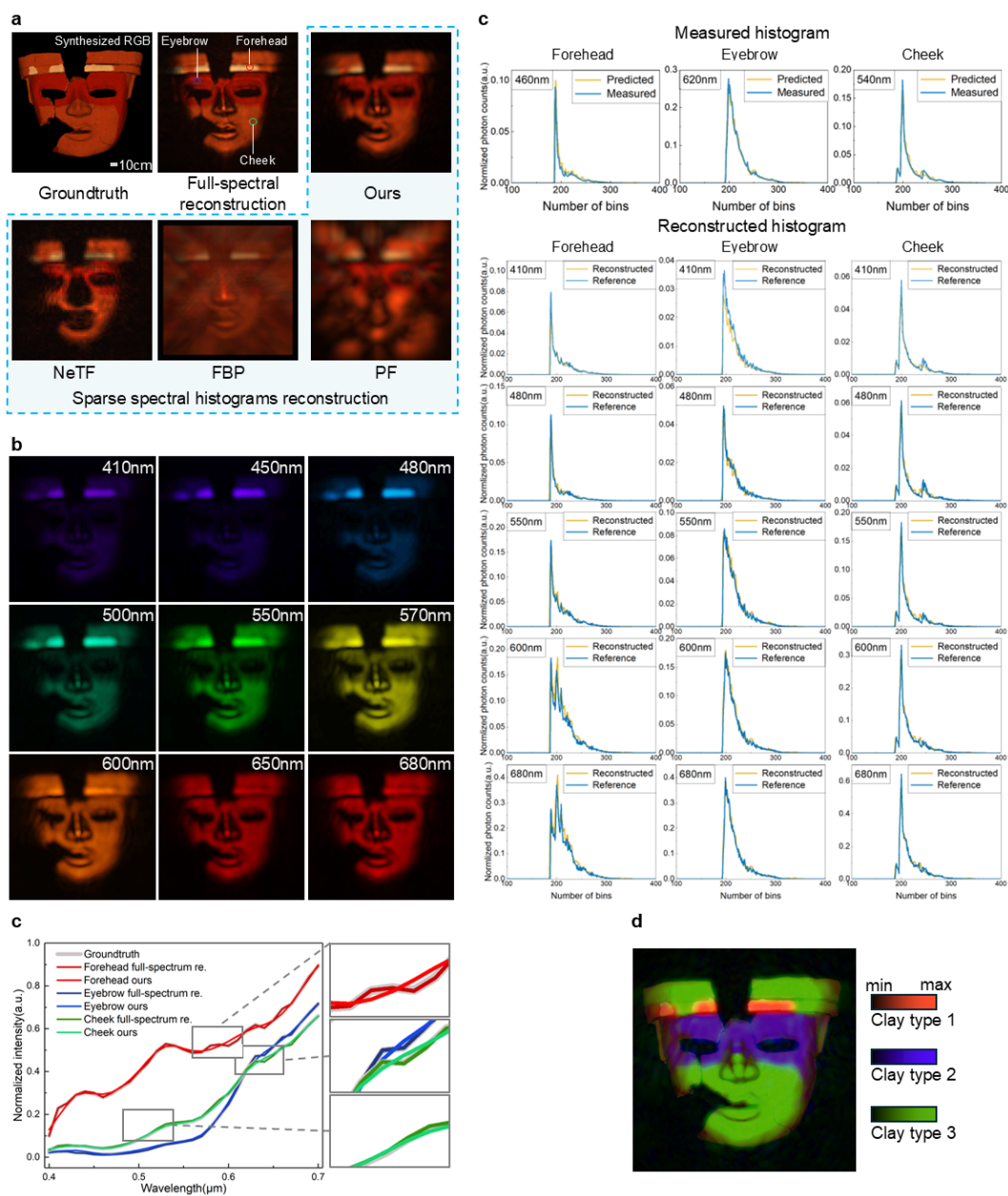


图 5-10 宽带反射率物体的重建

Figure 5-10 Evaluation of the proposed method on synthetic broadband natural spectral data.

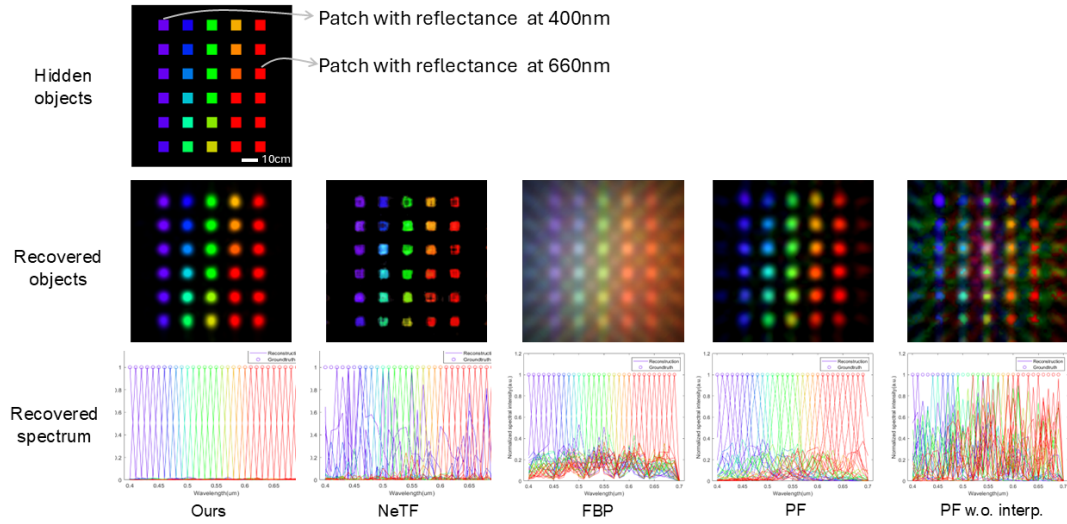


图 5-11 窄带反射贴片的重建

Figure 5-11 Reconstruction of synthetic narrow-band reflectance patches.

内容。BP 产生了具有病态反演特征的低对比度重建。

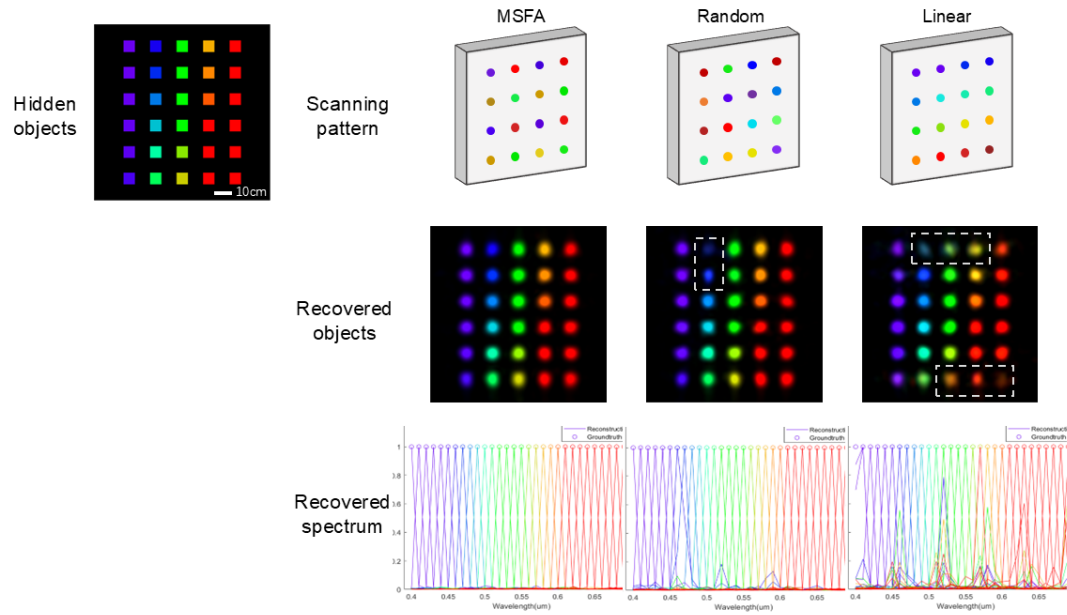


图 5-12 窄带反射贴片的重建

Figure 5-12 Reconstruction of synthetic narrow-band reflectance patches.

进一步对相同的贴片场景评估 MSFA、随机和线性扫描模式 (图 5-12)。MSFA 实现了准确的强度恢复和光谱对齐，而随机和线性模式显示了强度错误和光谱偏差。从定量上看，MSFA 达到了平均 PSNR 为 30.14 dB，优于随机 (29.91 dB) 和线性 (28.43 dB) 模式。

一个具有 32×32 扫描网格和 MSFA 模式的 'NLOS' 字母合成场景评估了宽

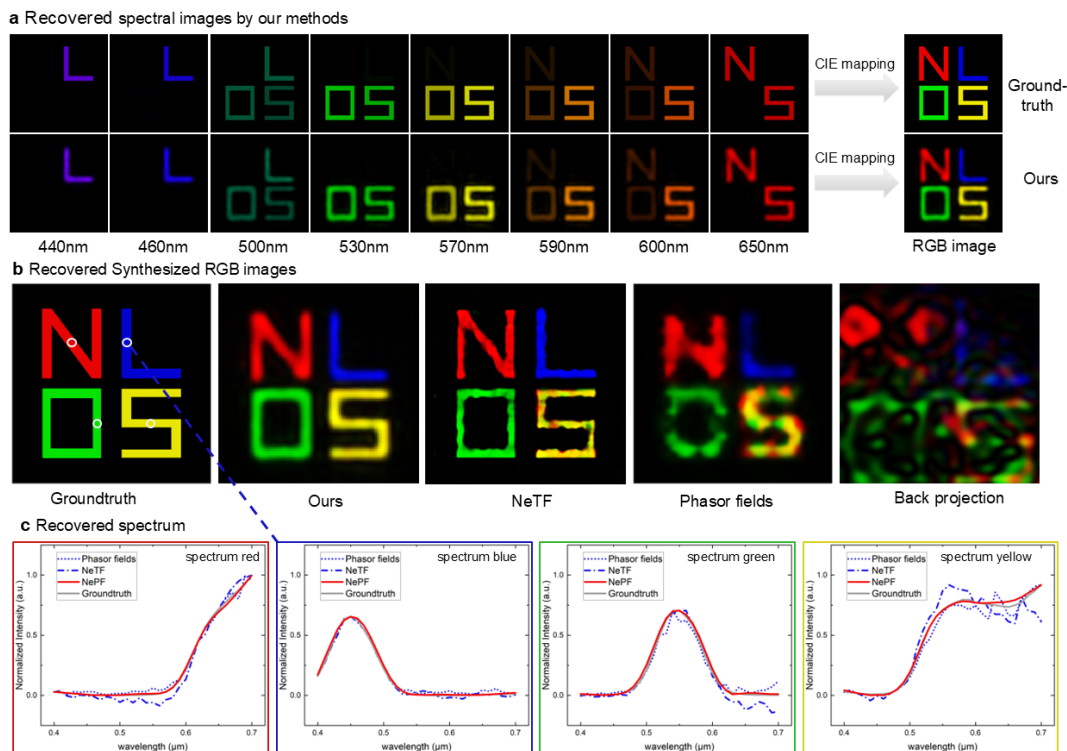


图 5-13 宽带'NLOS' 字母重建

Figure 5-13 Evaluation of the various method on synthetic broadband 'NLOS' letter data.

带性能 (图 5-13)。提出的方法重建了高保真度的光谱图像 (图 5-13a) 和准确的 RGB 映射 (图 5-13b), 与 NeTF/PF 的色彩伪影和 FBP 的失败模式形成对比。光谱剖面 (图 5-13c) 展示了提出的方法与地面真实情况的对齐 (PSNR 20.89dB), 优于其他替代方法。

图 5-14评估了在不同数据分辨率和噪声条件下的重建结果。数据分辨率为 32×32 , 64×64 , 128×128 , 带有泊松噪声的 128×128 , 以及带有时间抖动的 128×128 的恢复图像。在 32×32 分辨率下, 提出的方法保持了结构完整性 (PSNR 20.89 dB), 而 PF 引入了颜色失真 (17.21 dB) 并且 FBP 完全失败。在泊松噪声 (每像素 100 个光子) 和时间抖动 (170 ps 的高斯分布) 下, 提出的方法保持了边缘的锐度 (SSIM 0.65), 相比之下 NeTF (SSIM 0.58) 和 PF (SSIM 0.60)。

消融实验 (图 5-15) 揭示了光谱解码器和条件场的重要性。去除光谱解码器会引入色差, 而省略条件场则会降低空间一致性。在稀疏 (32×32) 采样下, 两个组件对于联合光谱-空间保真度都是至关重要的。

也在不同的材料上评估了提出的方法。如图 5-16所示, 模拟了四种不同材料的兔子数据: 木头、大理石、织物和金。木头、大理石、织物设置为漫反射, 而金设置为高度镜面反射。展示了提出的方法重建的几何形状和图像。提出的方法可以准确地重建木头、大理石和织物兔子的几何形状。虽然金兔子是镜面反射

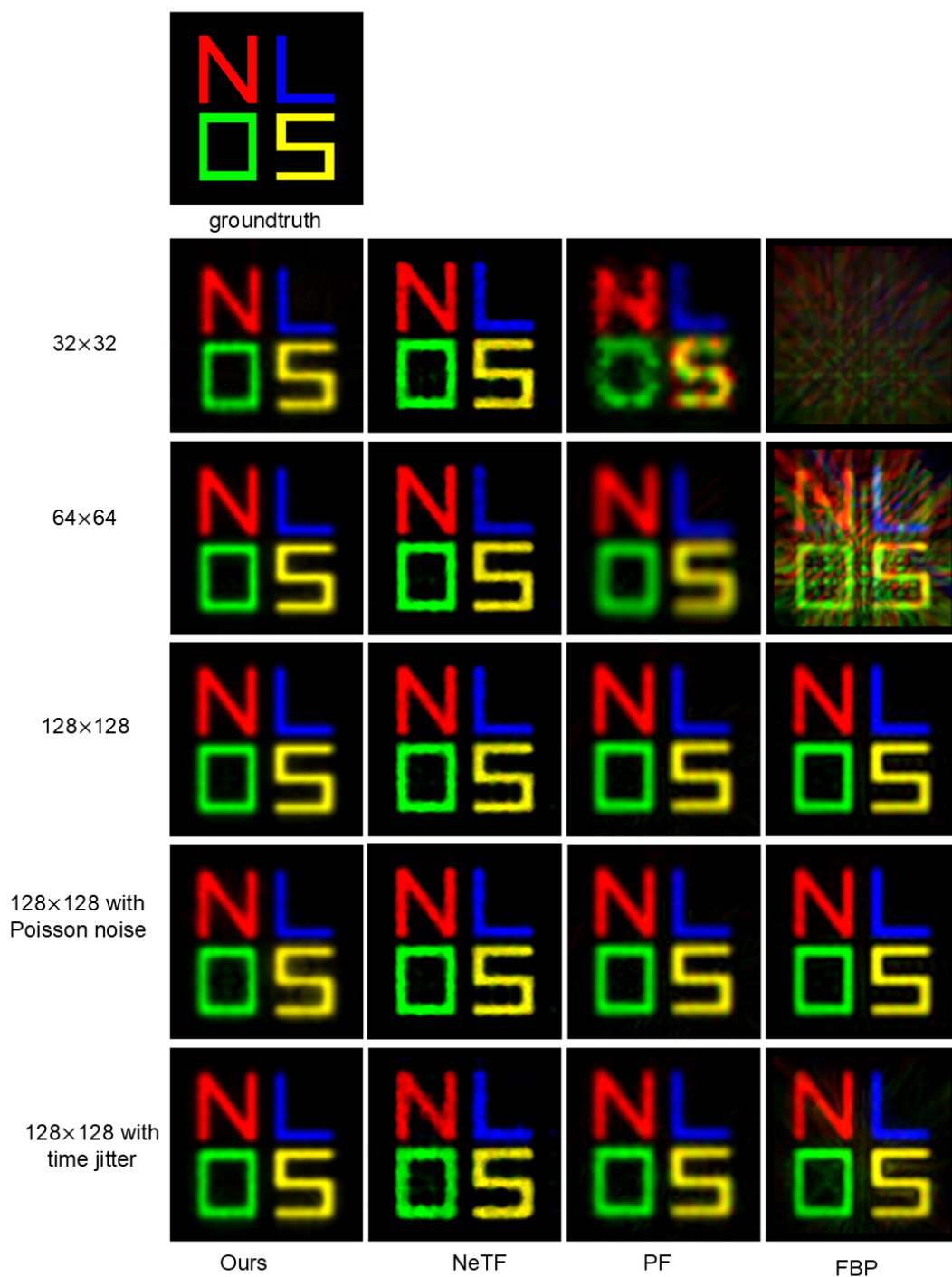


图 5-14 方法在不同的数据分辨率和噪声下的重建

Figure 5-14 Evaluate of the various method on data with different resolution and noise.

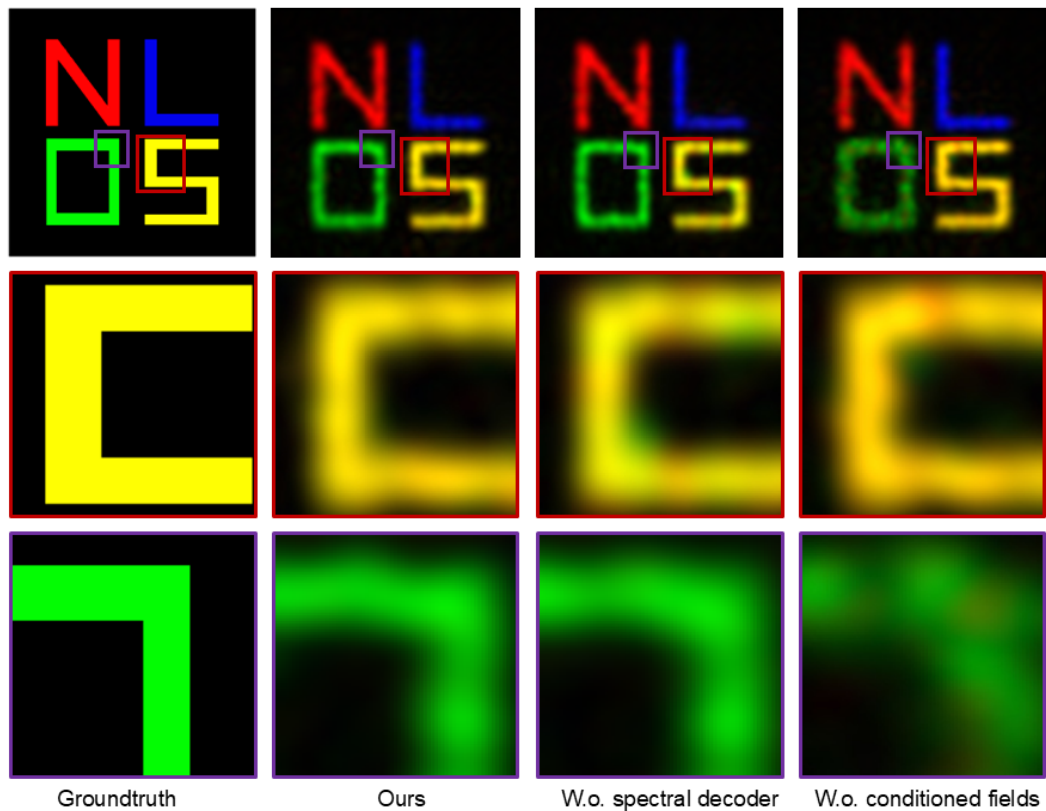


图 5-15 消融实验

Figure 5-15 Ablation study of our method.

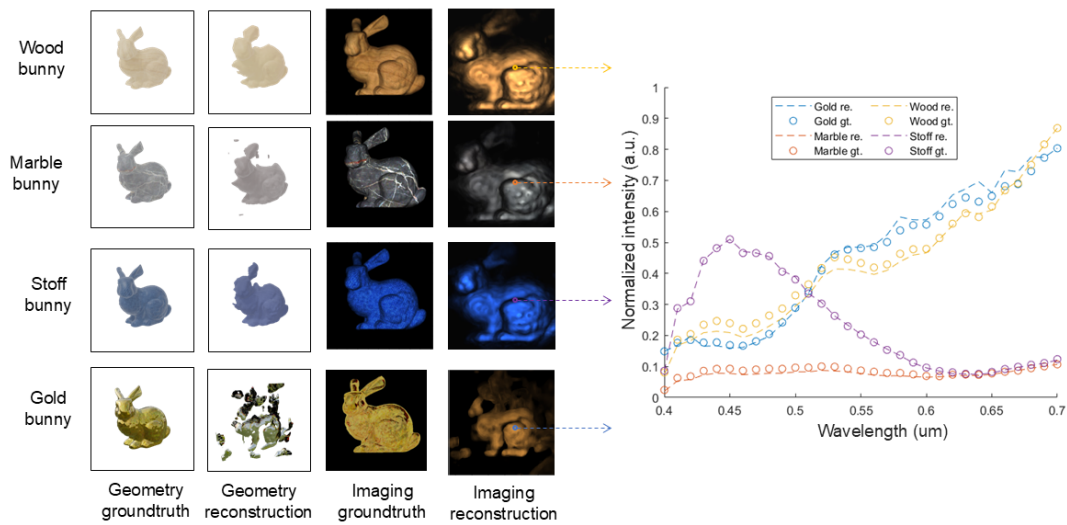


图 5-16 不同材质物体的重建

Figure 5-16 Reconstruction of different material synthetic bunny data.

的,但提出的方法只恢复了兔子的主要部分,并且有一些伪影。然而,提出的方法可以恢复所有材料制成的兔子的图像和光谱。这证明了提出的方法的有效性。

这些仿真数据上的重建结果确认了提出的方法对噪声的鲁棒性,对光谱复杂性的适应性,以及在实际成像条件下保持空间-光谱相关性上优于现有方法的优越性。

5.6.2 实测数据结果

进行实验,以评估设置和方法在各种隐藏场景中的表现。如图 5-17a 所示,首先关注以红色、蓝色、绿色和黄色呈现的字母“NLOS”。每个字母都是由反光材料制成,以增强从隐藏对象的光子反射。目标被放置在距可见中继墙 1 米的位置。在 1.6 米 $\times$ 1.6 米的正方形区域内,沿着一个 128 $\times$ 128 的网格进行光栅扫描,总曝光时间约为 4.5 小时。关注的光谱范围从 450 nm 到 750 nm,光谱分辨率为 10 nm,共有 31 个光谱带。所使用的扫描模式是多光谱滤波器阵列 (MSFA)。采用 4 ps 的时间分辨率,时间箱长度为 4096。图 5-17b 显示了原始的快照光谱瞬态,以及在 450 nm、500 nm、550 nm 和 600 nm 处的典型光子直方图。使用提出的方法、NeTF、PF 和 FBP 产生的恢复图像被合成成 RGB,以增强可视化。提出的方法有效地解决了不适定的重建问题和低信噪比 (SNR),而使用 NeTF、PF 和 FBP 获得的图像模糊且含有伪影。图 5-17a 中突出显示的选定像素的这些方法的光谱恢复,如图 5-17d 所示,证明了提出方法的光谱与实际值紧密相符。通过提出的方法恢复的光谱图像如图 5-17e 所示。

第三个实验分析了两片涂有不同反射材料的叶子 (图 5-18a)。实验设置与图 5-17 相同,只是扫描区域为 1 米 $\times$ 1 米。重建结果 (图 5-18b) 显示提出的方法准确地区分了材料类型,而比较的方法产生了噪音输出。光谱图 (图 5-18c) 在 500-550 nm 范围内显示出显著的差异,这也得到了光谱图像 (图 5-18d) 的证实。

进一步评估设置和方法在涉及真实荷花和假荷花的现实场景中的表现,如图 5-19 所示。隐藏对象的照片在图 5-19a 中呈现。实验设置与图 5-17 相似,只是扫描区域为 1 米 $\times$ 1 米。图 5-19b 显示了在 450 nm、550 nm、650 nm 和 750 nm 处的典型光子直方图,突出显示了数据的极低 SNR。提出的方法、NeTF、PF 和 FBP 恢复的图像被合成成 RGB,以提高可视化,如图 5-19c 所示。提出的方法熟练地处理了超低 SNR,并准确地区分了真实和假的荷花,而其他方法则表现出严重的伪影。图 5-19d 显示了提出的方法恢复的真实荷花和假荷花的光谱,突出显示了真实荷花在 700 nm 以上的波长,即近红外光谱中,反射率强,而假荷花则没有。通过提出的方法恢复的光谱图像也在图 5-19d 中显示。

荷叶的完整光谱数据重建在图 5-20 中描绘。所考虑的波长范围从 450nm 延伸到 690nm,光谱分辨率为 10nm。同时,时间箱从 160 延伸到 260,单位为 32ps,间隔为 20。在这个光谱范围内重构的瞬态表现出很高的连续性,从而验证了提出的方法的有效性。

最后,重建了一个几何复杂的马形象 (图 5-21, 左上角), 在 2 米 $\times$ 2 米的区域

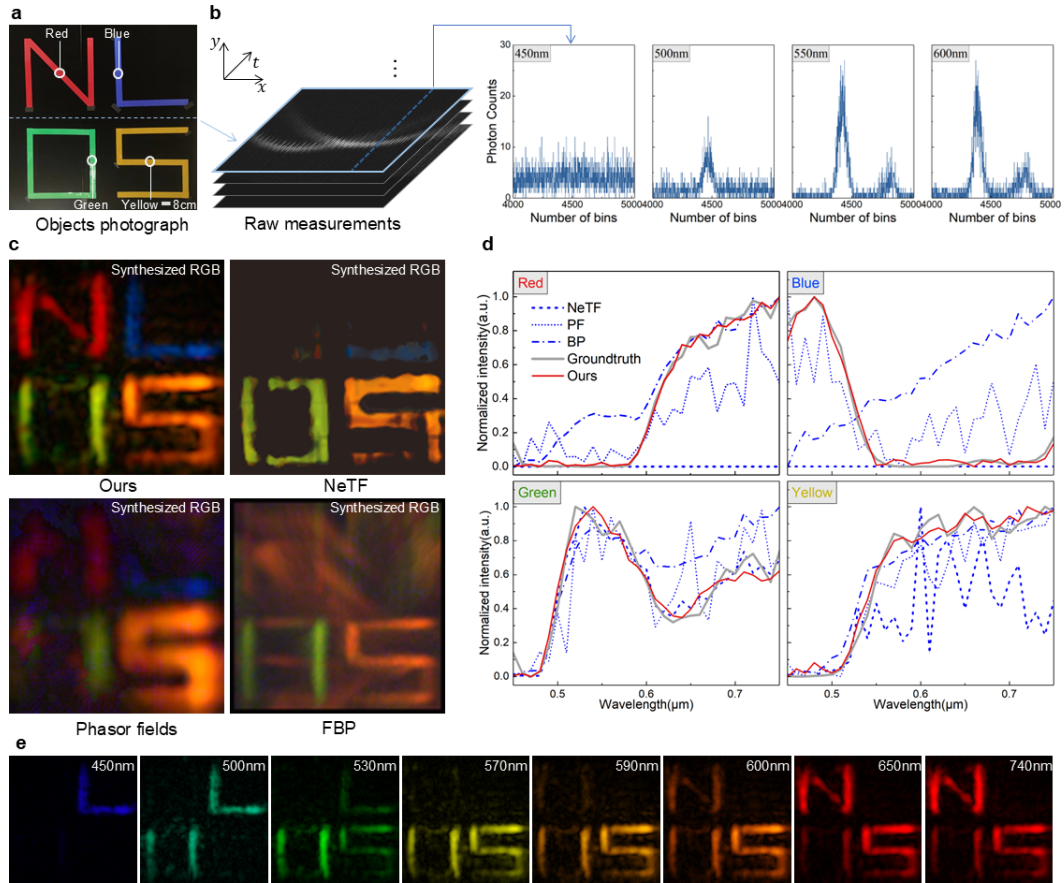


图 5-17 实测反光材质'NLOS'字母的重建

Figure 5-17 Evaluation of the proposed method on experimental data with retro-reflective hidden objects.

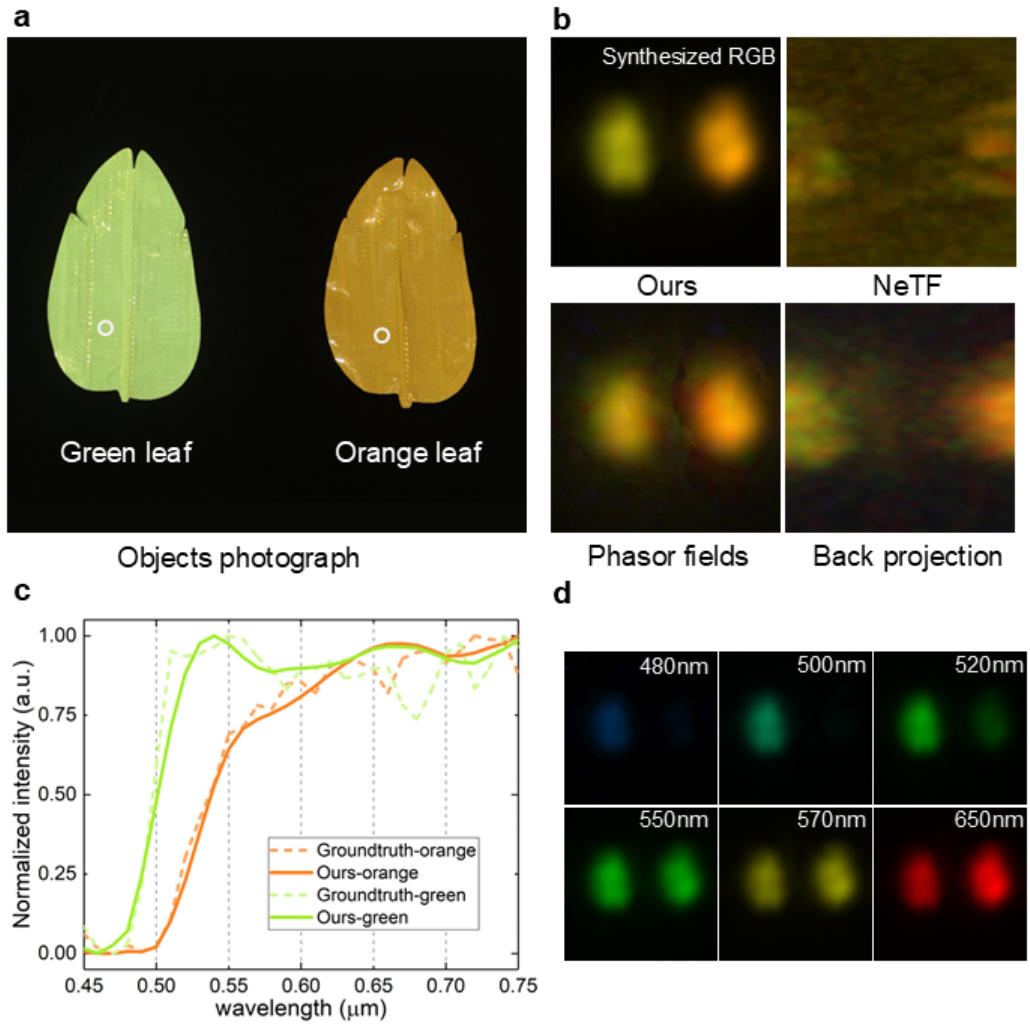


图 5-18 实测反光材质叶片的重建

Figure 5-18 Evaluation of the proposed method on experimental data with retro-reflective objects.

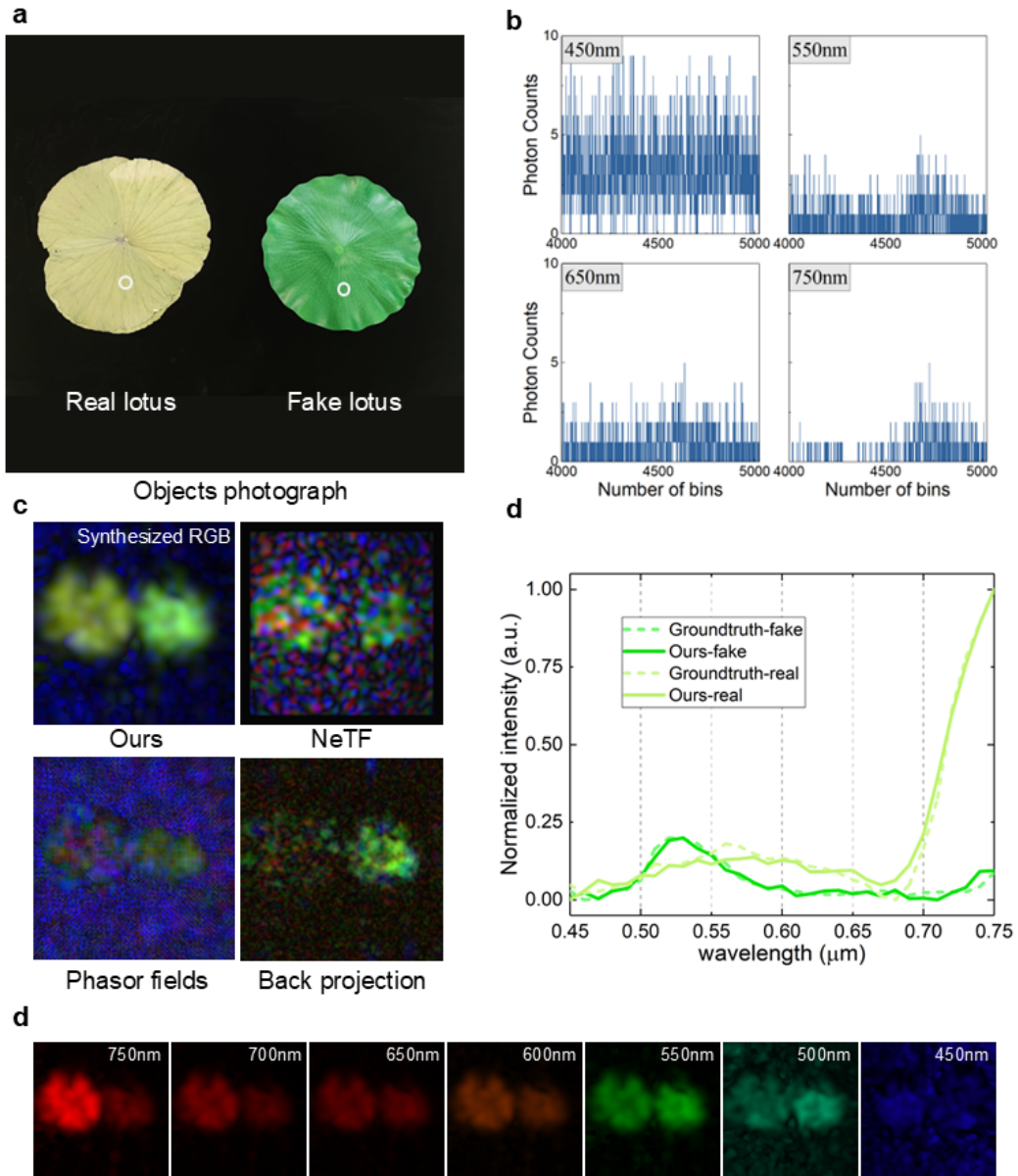


图 5-19 实测真假荷叶的重建

Figure 5-19 Evaluation of the proposed method on experimental data with natural objects.

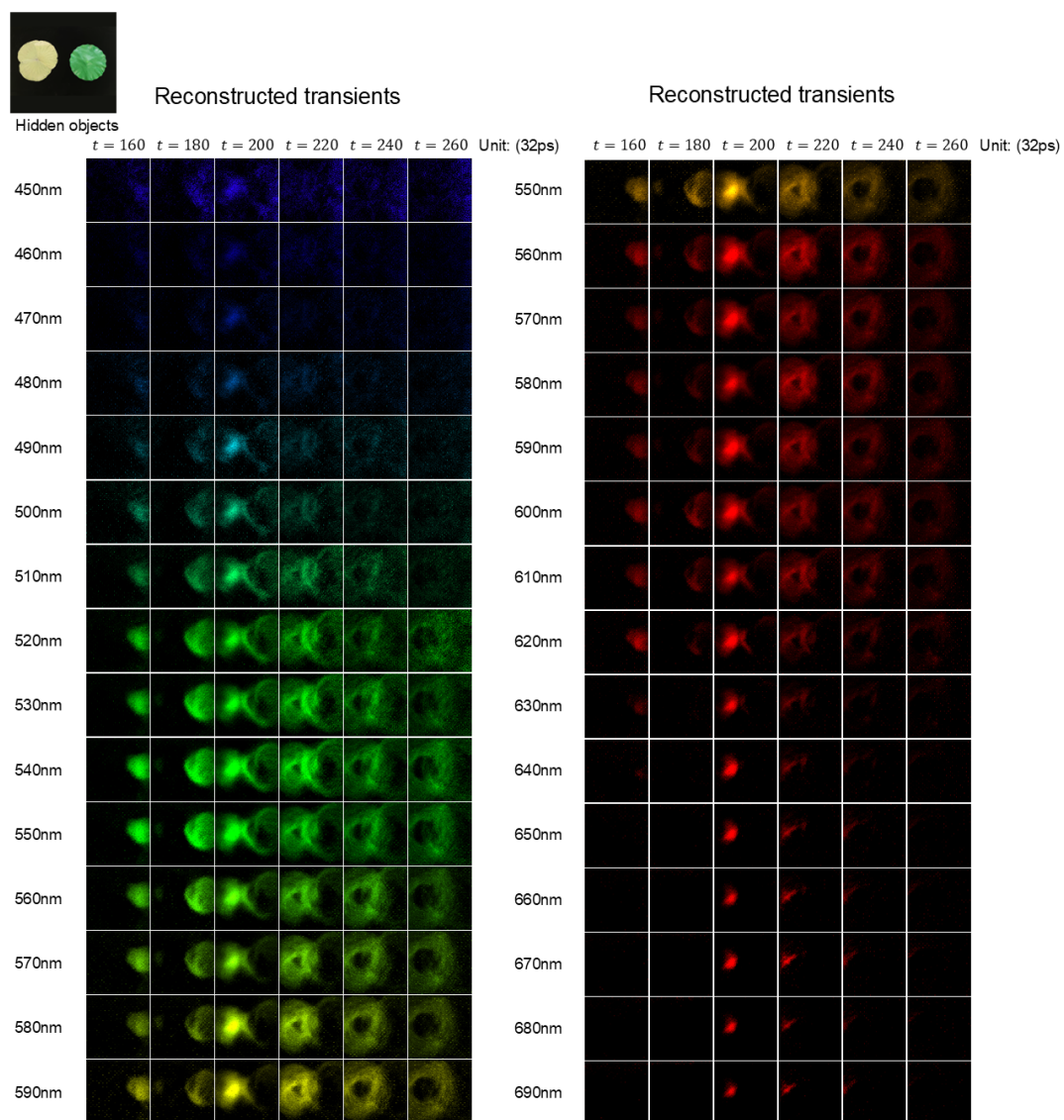


图 5-20 实测荷叶的完整光谱数据重建

Figure 5-20 Reconstructed spectral transients via our method on lotus scene.

内进行扫描。尽管信噪比极低，提出的方法仍然保留了细微的结构细节（图 5-21，右侧），而其他方法产生了伪影。与反向投影（BP）相比，恢复的光谱（图 5-21，左下角）进一步验证了提出的方法的准确性。

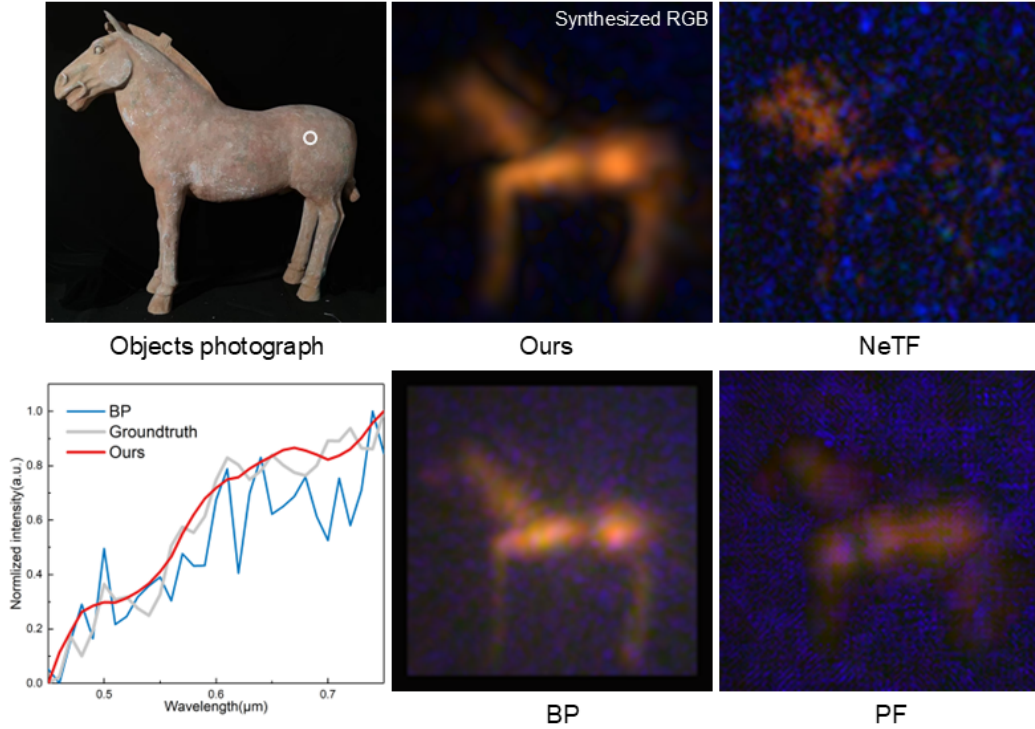


图 5-21 实测马的重建

Figure 5-21 Evaluation of the proposed method on experimental data with natural objects.

5.6.3 结论

提出了一个全面的快照光谱 NLOS 成像框架，该框架结合了波长模式化采集策略和 NeSPF，以解决采集效率、噪声鲁棒性和光谱保真度的关键挑战。通过设计一个 MSFA 扫描模式，提出的方法将采集时间和数据量减少到与单色 NLOS 成像相当的水平，同时实现高保真的 4D 光谱-空间重建。NeSPF 框架集成了相位场去噪，隐藏场景的神经隐式表示，和先验驱动的光谱解码，在窄带和宽带场景的合成和实验评估中，展示出优于现有方法的精度。与传统的 NeTF、PF 和 FBP 等方法相比，NeSPF 在保持高保真重建的同时，实现了高达 30 倍的采集速度提升，即使在超低信噪比条件下也能保持高保真度。该方法在解析材料特异性光谱特征方面的能力（例如，真实与合成荷叶的近红外反射）验证了其在需要无损光谱分析的应用中的潜力，包括考古检查^[160]、环境监测^[29]和生物医学诊断^[4]。

然而，也有几个限制需要考虑。首先，NeSPF 的计算需求导致每个场景的训练时间为 3 小时。其次，当前的光谱 NLOS 成像系统需要 3 小时的数据采集时间，这受到慢速扫描过程的限制。未来的工作将探索通过监督学习的神经场进

行实时推理，以加速光谱重建^[109,143,145,146]。此外，集成多光谱滤波器 SPAD 阵列可以提高光子采集效率，实现实时动态场景分析^[126,133,138]。尽管有这些挑战，提出方法的结果建立了光谱 NLOS 成像作为一种可扩展和实用的工具，推动了非视距环境计算成像的前沿。

第6章 结论与展望

本研究系统地探讨了神经网络在光谱成像领域的应用，重点关注了计算光谱成像的编码-解码框架，并分别提出了基于有监督学习、混合监督学习和预训练自监督学习的策略。

首先，我们提出了一种将物理仿真模型与有监督学习相结合的光谱超分辨率方法，有效地融合了光谱特征并学习了局部上下文信息。该方法已在长波红外多孔径快照式光谱成像领域得到了应用，与我们提出的气体辐射传输模型算法结合，证实了该技术在工业泄漏监测场景中的有效性。

其次，我们提出了一种基于混合监督学习的策略。这种策略首先利用物理模型模拟的数据进行有监督学习，然后通过神经光谱场优化神经网络输出的光谱曲线。最后，我们引入了基于物理的细化方法，以增强重构光谱与测量数据的一致性，从而实现高质量的光谱重建。在渐变带隙微型计算光谱仪上的实验结果证明，该方法在降低模型对标记数据依赖性的同时，能够达到较高的波长精度和光谱分辨率。

最后，我们进一步提出了一种基于预训练与自监督学习的框架，通过引导神经网络预训练优化光谱解码器，从而实现光谱信息的精确解析。同时，我们引入了光谱相位场，并将其纳入神经网络的训练中，以减小噪声干扰并提高光谱成像的质量。针对全局前向测量的光谱重建中的困难，我们的方法在非视距光谱成像中的应用，成功地实现了无需额外标记数据的端到端优化，大幅度降低了对实测数据量的需求。在仿真与实测验证中，该方法仅需传统方法 1/30 的数据量即可达到相同效果，为缓解训练样本不足问题提供了有效的解决方案。

总的来说，我们提出的神经网络光谱重建算法在各类光谱成像应用场景中都取得了优于以往的重建结果。

无论是有监督学习、混合监督学习还是自监督学习，都需要大量的数据样本支持网络训练。如果有足够的模拟或实验光谱数据，我们将有可能设计出更精准的神经网络结构。因此，开发能够产生大量高质量光谱数据的方法将是一个重要的研究方向。一种可能的方法是利用物理仿真模型和实验数据，结合深度学习生成对抗网络（GAN）生成高质量的仿真数据，以扩充训练样本。

另一个潜在的研究方向是多模态神经网络。目前公开的光谱数据集非常丰富，但其多样性导致这些数据不能通用，因此难以应用到同一个神经网络的训练中。如果能利用当前大模型框架中多模态训练的优势，整合多样性的高光谱数据，我们可能能够打造出通用光谱大模型，从而推动计算光谱成像领域的发展，打破各自为政的现状。

神经网络在光谱成像中的应用进程仍在发展阶段，我们期待有更多的研究者加入进来，利用神经网络的强大能力，推动光谱成像技术的发展。

参考文献

- [1] Kruse F A, Lefkoff A B, Boardman J W, et al. The spectral image processing system (sips) —interactive visualization and analysis of imaging spectrometer data [J]. Remote sensing of environment, 1993, 44(2-3): 145-163.
- [2] 王建宇. 成像光谱技术导论 [M]. Ke xue chu ban she, 2011.
- [3] Gat N, Subramanian S, Barhen J, et al. Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation, and manufacturing [C]//25th AIPR Workshop: Emerging Applications of Computer Vision: volume 2962. SPIE, 1997: 63-77.
- [4] Li Q, He X, Wang Y, et al. Review of spectral imaging technology in biomedical engineering: achievements and challenges [J]. Journal of biomedical optics, 2013, 18(10): 100901-100901.
- [5] Shimoni M, Haelterman R, Perneel C. Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques [J]. IEEE Geoscience and Remote Sensing Magazine, 2019, 7(2): 101-117.
- [6] Shaw G A, Burke H K. Spectral imaging for remote sensing [J]. Lincoln laboratory journal, 2003, 14(1): 3-28.
- [7] Barreto M A P, Johansen K, Angel Y, et al. Radiometric assessment of a uav-based push-broom hyperspectral camera [J]. Sensors, 2019, 19(21): 4699.
- [8] Favreau P, Hernandez C, Lindsey A S, et al. Thin-film tunable filters for hyperspectral fluorescence microscopy [J]. Journal of biomedical optics, 2014, 19(1): 011017-011017.
- [9] Hagen N, Kudenov M W. Review of snapshot spectral imaging technologies [J]. Optical Engineering, 2013, 52(9): 090901-090901.
- [10] Yang Z, Albrow-Owen T, Cai W, et al. Miniaturization of optical spectrometers [J]. Science, 2021, 371(6528): eabe0722.
- [11] Li Y, Xue Y, Tian L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media [J]. Optica, 2018, 5(10): 1181-1190.
- [12] Faccio D, Velten A, Wetzstein G. Non-line-of-sight imaging [J]. Nature Reviews Physics, 2020, 2(6): 318-327.
- [13] Bacca J, Martinez E, Arguello H. Computational spectral imaging: a contemporary overview [J]. Journal of the Optical Society of America A, 2023, 40(4): C115-C125.
- [14] Arce G R, Brady D J, Carin L, et al. Compressive coded aperture spectral imaging: An introduction [J]. IEEE Signal Processing Magazine, 2013, 31(1): 105-115.
- [15] Kittle D, Choi K, Wagadarikar A, et al. Multiframe image estimation for coded aperture snapshot spectral imagers [J]. Applied optics, 2010, 49(36): 6824-6833.
- [16] Hubold M, Montag E, Berlich R, et al. Multi-aperture system approach for snapshot multi-spectral imaging applications [J]. Optics Express, 2021, 29(5): 7361-7378.
- [17] Yoon H H, Fernandez H A, Nigmatulin F, et al. Miniaturized spectrometers with a tunable van der waals junction [J]. Science, 2022, 378(6617): 296-299.
- [18] Heide F, O' Toole M, Zang K, et al. Non-line-of-sight imaging with partial occluders and surface normals [J]. ACM Transactions on Graphics (ToG), 2019, 38(3): 1-10.
- [19] O' Toole M, Lindell D B, Wetzstein G. Confocal non-line-of-sight imaging based on the light-cone transform [J]. Nature, 2018, 555(7696): 338-341.

- [20] Chaitin G J. Algorithmic information theory [J]. IBM journal of research and development, 1977, 21(4): 350-359.
- [21] Roberts D A, Yaida S, Hanin B. The principles of deep learning theory: volume 46 [M]. Cambridge University Press Cambridge, MA, USA, 2022.
- [22] Yuan X, Brady D J, Katsaggelos A K. Snapshot compressive imaging: Theory, algorithms, and applications [J]. IEEE Signal Processing Magazine, 2021, 38(2): 65-88.
- [23] Huang L, Luo R, Liu X, et al. Spectral imaging with deep learning [J]. Light: Science & Applications, 2022, 11(1): 61.
- [24] Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report [J]. arXiv preprint arXiv:2303.08774, 2023.
- [25] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with alphafold 3 [J]. Nature, 2024, 630(8016): 493-500.
- [26] Zhao W X, Zhou K, Li J, et al. A survey of large language models [J]. arXiv preprint arXiv:2303.18223, 2023, 1(2).
- [27] Liu Y, Zhang K, Li Y, et al. Sora: A review on background, technology, limitations, and opportunities of large vision models [J]. arXiv preprint arXiv:2402.17177, 2024.
- [28] Wang Z, Yang Y, Yuan L, et al. Pixel-based long-wave infrared spectral image reconstruction using a hierarchical spectral transformer [J]. Sensors (Basel, Switzerland), 2024, 24(23): 7658.
- [29] Yang Y, Wang Z, Wang P, et al. Robust gas species and concentration monitoring via cross-talk transformer with snapshot infrared spectral imager [J]. Sensors and Actuators B: Chemical, 2024, 413: 135780.
- [30] Wang J, Pan B, Wang Z, et al. Single-pixel p-graded-n junction spectrometers [J]. Nature Communications, 2024, 15(1): 1773.
- [31] Shen S, Wang Z, Liu P, et al. Non-line-of-sight imaging via neural transient fields [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(7): 2257-2268.
- [32] Candès E J, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information [J]. IEEE Transactions on information theory, 2006, 52(2): 489-509.
- [33] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms [J]. Physica D: nonlinear phenomena, 1992, 60(1-4): 259-268.
- [34] Renard N, Bourennane S, Blanc-Talon J. Denoising and dimensionality reduction using multilinear tools for hyperspectral images [J]. IEEE Geoscience and Remote Sensing Letters, 2008, 5(2): 138-142.
- [35] Bioucas-Dias J M, Figueiredo M A. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration [J]. IEEE Transactions on Image processing, 2007, 16(12): 2992-3004.
- [36] Figueiredo M A, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems [J]. IEEE Journal of selected topics in signal processing, 2007, 1(4): 586-597.
- [37] Yuan X. Generalized alternating projection based total variation minimization for compressive sensing [C]//2016 IEEE International conference on image processing (ICIP). IEEE, 2016: 2539-2543.
- [38] Dong W, Shi G, Li X, et al. Compressive sensing via nonlocal low-rank regularization [J]. IEEE transactions on image processing, 2014, 23(8): 3618-3632.

- [39] Liu Y, Yuan X, Suo J, et al. Rank minimization for snapshot compressive imaging [J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(12): 2990-3006.
- [40] Yuan X, Tsai T H, Zhu R, et al. Compressive hyperspectral imaging with side information [J]. IEEE Journal of selected topics in Signal Processing, 2015, 9(6): 964-976.
- [41] Yang J, Liao X, Yuan X, et al. Compressive sensing by learning a gaussian mixture model from measurements [J]. IEEE Transactions on Image Processing, 2014, 24(1): 106-119.
- [42] Huang T, Yuan X, Dong W, et al. Deep gaussian scale mixture prior for image reconstruction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10778-10794.
- [43] Fu Y, Zhang T, Wang L, et al. Coded hyperspectral image reconstruction using deep external and internal learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7): 3404-3420.
- [44] Chen Y, Lai W, He W, et al. Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network [J]. IEEE Transactions on Image Processing, 2024, 33: 926-941.
- [45] Ma J, Liu X Y, Shou Z, et al. Deep tensor admn-net for snapshot compressive imaging [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 10223-10232.
- [46] Meng Z, Yu Z, Xu K, et al. Self-supervised neural networks for spectral snapshot compressive imaging [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 2622-2631.
- [47] Xie T, Liu L, Zhuang L. Plug-and-play priors for multi-shot compressive hyperspectral imaging [J]. IEEE Transactions on Image Processing, 2023, 32: 5326-5339.
- [48] Zheng S, Liu Y, Meng Z, et al. Deep plug-and-play priors for spectral snapshot compressive imaging [J]. Photonics Research, 2021, 9(2): B18-B29.
- [49] Chen Y, Gui X, Zeng J, et al. Combining low-rank and deep plug-and-play priors for snapshot compressive imaging [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023.
- [50] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [51] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [52] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [53] Jiang J, Sun H, Liu X, et al. Learning spatial-spectral prior for super-resolution of hyperspectral imagery [J]. IEEE Transactions on Computational Imaging, 2020, 6: 1082-1096.
- [54] Miao X, Yuan X, Pu Y, et al. l-net: Reconstruct hyperspectral images from a snapshot measurement [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4059-4069.
- [55] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1664-1673.

- [56] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution [C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, 2016: 694-711.
- [57] Shu C, Yu K, Duan Z, et al. Feature-metric loss for self-supervised learning of depth and egomotion [C]//European Conference on Computer Vision. Springer, 2020: 572-588.
- [58] Wang L, Wu Z, Zhong Y, et al. Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer [J]. *Photonics Research*, 2022, 10(8): 1848-1858.
- [59] Qin X, Quan Y, Ji H. Enhanced deep unrolling networks for snapshot compressive hyperspectral imaging [J]. *Neural Networks*, 2024, 174: 106250.
- [60] Meng Z, Yuan X, Jalali S. Deep unfolding for snapshot compressive imaging [J]. *International Journal of Computer Vision*, 2023, 131(11): 2933-2958.
- [61] Cheng Z, Chen B, Lu R, et al. Recurrent neural networks for snapshot compressive imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2264-2281.
- [62] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 9446-9454.
- [63] Sun Y, Yang Y, Liu Q, et al. Unsupervised spatial-spectral network learning for hyperspectral compressive snapshot reconstruction [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-14.
- [64] Bacca J, Fonseca Y, Arguello H. Compressive spectral image reconstruction using deep prior and low-rank tensor representation [J]. *Applied optics*, 2021, 60(14): 4197-4207.
- [65] Kim Y D, Choi S. Nonnegative tucker decomposition [C]//2007 IEEE conference on computer vision and pattern recognition. IEEE, 2007: 1-8.
- [66] Li M, Fu Y, Zhang T, et al. Supervise-assisted self-supervised deep-learning method for hyperspectral image restoration [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.
- [67] Candes E J. The restricted isometry property and its implications for compressed sensing [J]. *Comptes rendus. Mathematique*, 2008, 346(9-10): 589-592.
- [68] Arguello H, Arce G R. Colored coded aperture design by concentration of measure in compressive spectral imaging [J]. *IEEE Transactions on Image Processing*, 2014, 23(4): 1896-1908.
- [69] Wang L, Zhang T, Fu Y, et al. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging [J]. *IEEE Transactions on Image Processing*, 2018, 28(5): 2257-2270.
- [70] Galvis L, Lau D, Ma X, et al. Coded aperture design in compressive spectral imaging based on side information [J]. *Applied optics*, 2017, 56(22): 6332-6340.
- [71] Lin X, Wetzstein G, Liu Y, et al. Dual-coded compressive hyperspectral imaging [J]. *Optics letters*, 2014, 39(7): 2044-2047.
- [72] Shi Z, Dun X, Wei H, et al. Learned multi-aperture color-coded optics for snapshot hyperspectral imaging [J]. *ACM Transactions on Graphics (TOG)*, 2024, 43(6): 1-11.
- [73] Hinojosa C, Bacca J, Arguello H. Coded aperture design for compressive spectral subspace clustering [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2018, 12(6): 1589-1600.
- [74] Courbariaux M, Bengio Y, David J P. Binaryconnect: Training deep neural networks with binary weights during propagations [J]. *Advances in neural information processing systems*, 2015, 28.

- [75] Zhang W, Song H, He X, et al. Deeply learned broadband encoding stochastic hyperspectral imaging [J]. *Light: Science & Applications*, 2021, 10(1): 108.
- [76] Jia Y, Zheng Y, Gu L, et al. From rgb to spectrum for natural scenes via manifold-based mapping [C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 4705-4713.
- [77] Tenenbaum J B, Silva V d, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. *science*, 2000, 290(5500): 2319-2323.
- [78] Hirai A, Inoue T, Itoh K, et al. Application of measurement multiple-image fourier of fast phenomena transform spectral imaging to measurement of fast phenomena [J]. *Optical Review*, 1994, 1: 205-207.
- [79] Kudenov M W, Dereniak E L. Compact real-time birefringent imaging spectrometer [J]. *Optics express*, 2012, 20(16): 17973-17986.
- [80] Hubold M, Berlich R, Gassner C, et al. Ultra-compact micro-optical system for multispectral imaging [C]//*MOEMS and Miniaturized Systems XVII: volume 10545*. SPIE, 2018: 206-213.
- [81] Mu T, Han F, Bao D, et al. Compact snapshot optically replicating and remapping imaging spectrometer (orris) using a focal plane continuous variable filter [J]. *Optics Letters*, 2019, 44(5): 1281-1284.
- [82] Yang Y, Liu S, Wang P, et al. Uncooled snapshot infrared spectrometer with improved sensitivity for gas imaging [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 73: 1-9.
- [83] Gittins C M. Detection and characterization of chemical vapor fugitive emissions by nonlinear optimal estimation: theory and simulation [J]. *Applied Optics*, 2009, 48(23): 4545-4561.
- [84] Manley M. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials [J]. *Chemical Society Reviews*, 2014, 43(24): 8200-8214.
- [85] Ryckeboer E, Bockstaele R, Vanslembrouck M, et al. Glucose sensing by waveguide-based absorption spectroscopy on a silicon chip [J]. *Biomedical optics express*, 2014, 5(5): 1636-1648.
- [86] Crocombe R A. Portable spectroscopy [J]. *Applied spectroscopy*, 2018, 72(12): 1701-1751.
- [87] Cheng R, Zou C L, Guo X, et al. Broadband on-chip single-photon spectrometer [J]. *Nature communications*, 2019, 10(1): 4104.
- [88] Yang Z, Albrow-Owen T, Cui H, et al. Single-nanowire spectrometers [J]. *Science*, 2019, 365(6457): 1017-1020.
- [89] Goldman D S, White P, Anheier N. Miniaturized spectrometer employing planar waveguides and grating couplers for chemical analysis [J]. *Applied optics*, 1990, 29(31): 4583-4589.
- [90] Subramanian A Z, Ryckeboer E, Dhakal A, et al. Silicon and silicon nitride photonic circuits for spectroscopic sensing on-a-chip [J]. *Photonics Research*, 2015, 3(5): B47-B59.
- [91] Sander D, Müller J. Selffocussing phase transmission grating for an integrated optical microspectrometer [J]. *Sensors and Actuators A: Physical*, 2001, 88(1): 1-9.
- [92] Cheben P, Schmid J, Delâge A, et al. A high-resolution silicon-on-insulator arrayed waveguide grating microspectrometer with sub-micrometer aperture waveguides [J]. *Optics express*, 2007, 15(5): 2299-2306.
- [93] Gao B, Shi Z, Boyd R W. Design of flat-band superprism structures for on-chip spectroscopy [J]. *Optics express*, 2015, 23(5): 6491-6496.

- [94] Calafiore G, Koshelev A, Dhuey S, et al. Holographic planar lightwave circuit for on-chip spectroscopy [J]. *Light: Science & Applications*, 2014, 3(9): e203-e203.
- [95] Zhang H, Wang X L, Soos J I, et al. Design of a miniature solid state nir spectrometer [C]// *Infrared Detectors and Instrumentation for Astronomy: volume 2475*. SPIE, 1995: 376-383.
- [96] Gat N. Imaging spectroscopy using tunable filters: a review [J]. *Wavelet Applications VII*, 2000, 4056: 50-64.
- [97] Mallinson S, Jerman J. Miniature micromachined fabry-perot interferometers in silicon [J]. *Electronics letters*, 1987, 23(20): 1041-1043.
- [98] Carmo J P, Rocha R P, Bartek M, et al. A review of visible-range fabry-perot microspectrometers in silicon for the industry [J]. *Optics & Laser Technology*, 2012, 44(7): 2312-2320.
- [99] Antila J, Miranto A, Mäkynen J, et al. Mems and piezo actuator-based fabry-perot interferometer technologies and applications at vtt [C]// *Next-Generation Spectroscopic Technologies III: volume 7680*. SPIE, 2010: 192-203.
- [100] Nitkowski A, Chen L, Lipson M. Cavity-enhanced on-chip absorption spectroscopy using microring resonators [J]. *Optics express*, 2008, 16(16): 11930-11936.
- [101] Zavracky P M, Hennenberg E. Miniature fabry perot spectrometers using micromachining technology [C]// *Proceedings of WESCON'95*. IEEE, 1995: 325.
- [102] Ebermann M, Neumann N, Hiller K, et al. Tunable mems fabry-pérot filters for infrared microspectrometers: a review [J]. *MOEMS and miniaturized systems XV*, 2016, 9760: 64-83.
- [103] Gunning W, Yeh P. Multiple-cavity infrared electro-optic tunable filter [C]// *Active optical devices: volume 202*. SPIE, 1980: 21-25.
- [104] Yao Y, Hou J, Liu H, et al. Design of programmable multi-wavelength tunable filter on lithium niobate [J]. *Results in Physics*, 2019, 15: 102741.
- [105] Gunning W, Pasko J, Tracy J. A liquid crystal tunable spectral filter: visible and infrared operation [C]// *Imaging Spectroscopy I: volume 268*. SPIE, 1981: 190-194.
- [106] Lammel G, Schweizer S, Renaud P. Microspectrometer based on a tunable optical filter of porous silicon [J]. *Sensors and Actuators A: Physical*, 2001, 92(1-3): 52-59.
- [107] Liao J H, Wang W, Chen C J, et al. Design and fabrication of large-area tunable moems-based shortwave infrared fabry-pérot filters [J]. *Journal of Vacuum Science & Technology B*, 2019, 37(3).
- [108] Arad B, Ben-Shahar O. Sparse recovery of hyperspectral signal from natural rgb images [C]// *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VII 14*. Springer, 2016: 19-34.
- [109] Chen W, Wei F, Kutulakos K N, et al. Learned feature embeddings for non-line-of-sight imaging and recognition [J]. *ACM Transactions on Graphics (ToG)*, 2020, 39(6): 1-18.
- [110] Bouman K L, Ye V, Yedidia A B, et al. Turning corners into cameras: Principles and methods [C]// *Proceedings of the IEEE International Conference on Computer Vision*. 2017: 2270-2278.
- [111] Tancik M, Swedish T, Satat G, et al. Data-driven non-line-of-sight imaging with a traditional camera [C]// *Imaging Systems and Applications*. Optical Society of America, 2018: IW2B-6.
- [112] Saunders C, Murray-Bruce J, Goyal V K. Computational periscopy with an ordinary digital camera [J]. *Nature*, 2019, 565(7740): 472-475.

- [113] Tanaka K, Mukaigawa Y, Kadambi A. Polarized non-line-of-sight imaging [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2136-2145.
- [114] Sharma P, Aittala M, Schechner Y Y, et al. What you can learn by staring at a blank wall [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2330-2339.
- [115] Musarra G, Lyons A, Conca E, et al. Non-line-of-sight three-dimensional imaging with a single-pixel camera [J]. *Physical Review Applied*, 2019, 12(1): 011002.
- [116] Ahn B, Dave A, Veeraraghavan A, et al. Convolutional approximations to the general non-line-of-sight imaging operator [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 7889-7899.
- [117] Buttafava M, Zeman J, Tosi A, et al. Non-line-of-sight imaging using a time-gated single photon avalanche diode [J]. *Optics express*, 2015, 23(16): 20997-21011.
- [118] Liu X, Wang J, Xiao L, et al. Non-line-of-sight imaging with arbitrary illumination and detection pattern [J]. *Nature Communications*, 2023, 14(1): 3230.
- [119] O' Toole M, Lindell D B, Wetzstein G. Confocal non-line-of-sight imaging based on the light-cone transform [J]. *Nature*, 2018, 555(7696): 338-341.
- [120] Lindell D B, Wetzstein G, O'Toole M. Wave-based non-line-of-sight imaging using fast fk migration [J]. *ACM Transactions on Graphics (TOG)*, 2019, 38(4): 1-13.
- [121] Liu X, Guillén I, La Manna M, et al. Non-line-of-sight imaging using phasor-field virtual wave optics [J]. *Nature*, 2019, 572(7771): 620-623.
- [122] Katz O, Heidmann P, Fink M, et al. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations [J]. *Nature photonics*, 2014, 8(10): 784-790.
- [123] Cao R, De Goumoëns F, Blochet B, et al. High-resolution non-line-of-sight imaging employing active focusing [J]. *Nature Photonics*, 2022, 16(6): 462-468.
- [124] Katz O, Small E, Silberberg Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light [J]. *Nature photonics*, 2012, 6(8): 549-553.
- [125] Ye J T, Sun Y, Li W, et al. Real-time non-line-of-sight computational imaging using spectrum filtering and motion compensation [J]. *Nature Computational Science*, 2024: 1-8.
- [126] Nam J H, Brandt E, Bauer S, et al. Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second [J]. *Nature communications*, 2021, 12(1): 6526.
- [127] Willomitzer F, Rangarajan P V, Li F, et al. Fast non-line-of-sight imaging with high-resolution and wide field of view using synthetic wavelength holography [J]. *Nature communications*, 2021, 12(1): 6647.
- [128] Czajkowski R, Murray-Bruce J. Two-edge-resolved three-dimensional non-line-of-sight imaging with an ordinary camera [J]. *Nature Communications*, 2024, 15(1): 1162.
- [129] Metzler C A, Lindell D B, Wetzstein G. Keyhole imaging: non-line-of-sight imaging and tracking of moving objects along a single optical path [J]. *IEEE Transactions on Computational Imaging*, 2020, 7: 1-12.
- [130] Jiang D, Liu X, Luo J, et al. Ring and radius sampling based phasor field diffraction algorithm for non-line-of-sight reconstruction [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(11): 7841-7853.
- [131] Isogawa M, Chan D, Yuan Y, et al. Efficient non-line-of-sight imaging from transient sinograms [C]//European conference on computer vision. Springer, 2020: 193-208.

- [132] Li Y, Zhang Y, Ye J, et al. Deep non-line-of-sight imaging from under-scanning measurements [J]. *Advances in Neural Information Processing Systems*, 2023, 36: 59095-59106.
- [133] Seidel S, Rueda-Chacón H, Cusini I, et al. Non-line-of-sight snapshots and background mapping with an active corner camera [J]. *Nature Communications*, 2023, 14(1): 3677.
- [134] La Manna M, Nam J H, Azer Reza S, et al. Non-line-of-sight-imaging using dynamic relay surfaces [J]. *Optics express*, 2020, 28(4): 5331-5339.
- [135] Ye J T, Huang X, Li Z P, et al. Compressed sensing for active non-line-of-sight imaging [J]. *Optics Express*, 2021, 29(2): 1749-1763.
- [136] Wu C, Liu J, Huang X, et al. Non-line-of-sight imaging over 1.43 km [J]. *Proceedings of the National Academy of Sciences*, 2021, 118(10): e2024468118.
- [137] Ye J, Hong Y, Su X, et al. Plug-and-play algorithms for dynamic non-line-of-sight imaging [J]. *ACM Transactions on Graphics*, 2024, 43(5): 1-12.
- [138] Pei C, Zhang A, Deng Y, et al. Dynamic non-line-of-sight imaging system based on the optimization of point spread functions [J]. *Optics Express*, 2021, 29(20): 32349-32364.
- [139] Wu D, Wetzstein G, Barsi C, et al. Frequency analysis of transient light transport with applications in bare sensor imaging [C]//*Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7-13, 2012, *Proceedings, Part I* 12. Springer, 2012: 542-555.
- [140] Thrampoulidis C, Shulkind G, Xu F, et al. Exploiting occlusion in non-line-of-sight active imaging [J]. *IEEE Transactions on Computational Imaging*, 2018, 4(3): 419-431.
- [141] Lindell D B, Wetzstein G, O'Toole M. Wave-based non-line-of-sight imaging using fast fk migration [J]. *ACM Transactions on Graphics (ToG)*, 2019, 38(4): 1-13.
- [142] Liu X, Velten A. The role of wigner distribution function in non-line-of-sight imaging [C]//*2020 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2020: 1-12.
- [143] Yu Y, Shen S, Wang Z, et al. Enhancing non-line-of-sight imaging via learnable inverse kernel and attention mechanisms [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023: 10563-10573.
- [144] Galindo M, Marco J, O'Toole M, et al. A dataset for benchmarking time-resolved non-line-of-sight imaging [M]//*ACM SIGGRAPH 2019 Posters*. 2019: 1-2.
- [145] Mu F, Mo S, Peng J, et al. Physics to the rescue: Deep non-line-of-sight reconstruction for high-speed imaging [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [146] Li Y, Peng J, Ye J, et al. Nlost: Non-line-of-sight imaging with transformer [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 13313-13322.
- [147] Liu X, Bauer S, Velten A. Analysis of feature visibility in non-line-of-sight measurements [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 10140-10148.
- [148] Pediredla A, Dave A, Veeraraghavan A. Snlos: Non-line-of-sight scanning through temporal focusing [C]//*2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019: 1-13.
- [149] Grau Chopite J, Hullin M B, Wand M, et al. Deep non-line-of-sight reconstruction [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 960-969.

- [150] Arellano V, Gutierrez D, Jarabo A. Fast back-projection for non-line of sight reconstruction [M]//ACM SIGGRAPH 2017 Posters. 2017: 1-2.
- [151] La Manna M, Kine F, Breitbach E, et al. Error backprojection algorithms for non-line-of-sight imaging [J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(7): 1615-1626.
- [152] Velten A, Willwacher T, Gupta O, et al. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging [J]. Nature communications, 2012, 3(1): 745.
- [153] Young S I, Lindell D B, Girod B, et al. Non-line-of-sight surface reconstruction using the directional light-cone transform [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1407-1416.
- [154] Liu X, Guillén I, La Manna M, et al. Non-line-of-sight imaging using phasor-field virtual wave optics [J]. Nature, 2019, 572(7771): 620-623.
- [155] Liu X, Bauer S, Velten A. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems [J]. Nature communications, 2020, 11(1): 1645.
- [156] Liu X, Wang J, Li Z, et al. Non-line-of-sight reconstruction with signal-object collaborative regularization [J]. Light: Science & Applications, 2021, 10(1): 198.
- [157] Shen S, Wang Z, Liu P, et al. Non-line-of-sight imaging via neural transient fields [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(7): 2257-2268.
- [158] Fujimura Y, Kushida T, Funatomi T, et al. Nlos-neus: Non-line-of-sight neural implicit surface [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 10532-10541.
- [159] Miao L, Qi H. The design and evaluation of a generic method for generating mosaicked multispectral filter arrays [J]. IEEE Transactions on Image Processing, 2006, 15(9): 2780-2791.
- [160] Conrey R M, Goodman-Elgar M, Bettencourt N, et al. Calibration of a portable x-ray fluorescence spectrometer in the analysis of archaeological samples using influence coefficients [J]. Geochemistry: Exploration, Environment, Analysis, 2014, 14(3): 291-301.

致 谢

在上海科技大学和上海技术物理研究所总共六年的学习与研究生活，对我的人生产生了巨大的影响。上科大和技物所宽松活泼的学习与讨论氛围，为我的学习与发展提供了宝贵的平台，保障了优越的科研条件。老师与同窗们之间的交流，总能碰撞出灵感的火花。因此，谨向那些在我的工作学习和生活中，为我提供帮助与交流的人，致以衷心的感谢。

首先，我特别感谢我的导师，王建宇老师。王老师严谨的治学态度，广博的学识，以及对工程实践的严格要求，对我学术发展方向的自由宽容，对我产生了重要的影响，是我过去和未来学习的榜样。我在读博期间与多位不同方向不同单位的老师进行深度合作，与王老师的支持和鼓励密不可分。

我还要特别感谢我在技物所的第二位导师，李春来老师。李春来老师继承了王建宇老师的宽容开放的学术态度和严谨的工程要求，我有幸向李老师学习多年，深深被李老师全身心投入科研、对学生宽容开放的思想所感动。

我要特别感谢我在上科大最亲密最敬仰的导师和合作者虞晶怡老师。在虞老师的指导下，我学习到了科研工作的基本方式，同时也在虞老师的引导下学到了许多做人做事的方式。我在上科大期间，在虞老师的引荐下我能够有机会与多位不同领域的老师合作，见识领悟到不同领域的科研风格，也塑造了我自己的科研方式。

我还要感谢一起合作的李实英老师，与学生一起进行长期的科研攻关，指导我们的科研方向。感谢一起合作的陈佰乐老师，任无畏老师。感谢曾指导过的王雄老师、马月昕老师、陈文拯老师、陆卫老师、王少伟老师。

感谢技物所的袁立银老师，谢佳楠师兄、金健师兄、金海军师兄、刘馨泽、孔垂丰、张玉琛。

感谢技物所刘世界师兄、唐国良师兄、王旭师兄、黄骏师兄、王梅竹师姐、赵海艇师兄、姬忠鹏师兄、李津宇师兄、吴兵师兄、徐艳师姐、玄志一师兄、刘清权师兄。

感谢技物所的杨扬、文天成、徐佳飞、陈少毅、马天镇、司岳、谭永健、何智静、王森远、程博、李想、王朋宇、马瑶、孔艳洁、刘倩敏、杨诗承。

感谢上科大的吴旻烨师兄、沈思远、王景熠、潘贝贝、潘正卿、李睿潜、王悦涵、余焱华、刘平、高天、夏苏安、黄彬彬、刘洋、彭星月、王子衡、金奕彬、张家恺、李林霖、奚梓钧、邓智城、张圣伟、张启煊、韩嘉恒、姚凯忻、李敏章、蒋颖文琪、王有佳、张翰宇、张钰豪、宋承宣。

感谢我的父母多年以来对我的辛勤培养，对我发展方向的理解与支持，为我解除了后顾之忧。

再次感谢所有关心支持我的人，谢谢大家。

2025 年 6 月

作者简历及攻读学位期间发表的学术论文与其他相关学术成果

作者简历：

2015 年 9 月——2019 年 6 月，在北京航空航天大学电子信息工程学院获得学士学位。

2019 年 9 月——2025 年 6 月，在中国科学院上海技术物理研究所和上海科技大学攻读博士学位。

已发表（或正式接受）的学术论文：

- (1) Jingyi Wang[†], Beibei Pan[†], **Zi Wang**[†], Jiakai Zhang, Zhiqi Zhou, Lu Yao, Yanan Wu, Wuwei Ren, Jianyu Wang, Haiming Ji, Jingyi Yu* and Baile Chen*. Single-pixel p-graded-n junction spectrometers. *Nature Communications*, 15(1), 1773. 2024.
- (2) Yang Yang[†], **Zi Wang**[†], Pengyu Wang, Guoliang Tang, Chengyu Liu, Chunlai* Li, Jianyu Wang*. Robust gas species and concentration monitoring via cross-talk transformer with snapshot infrared spectral imager. *Sensors and Actuators B: Chemical*, 413, 135780. 2024.
- (3) Siyuan Shen[†], **Zi Wang**[†], Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiy-ing Li* and Jingyi Yu*. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7), 2257-2268. 2021.
- (4) **Zi Wang**, Yang Yang, Liyin Yuan, Chunlai Li*, Jianyu Wang*. Pixel-Based Long-Wave Infrared Spectral Image Reconstruction Using a Hierarchical Spectral Transformer. *Sensors (Basel, Switzerland)*, 24(23), 7658. 2024.
- (5) Zhiyi Xuan, **Zi Wang**, Qingquan Liu, Songlei Huang, Bo Yang, Liyi Yang, Zhiqin Yin, Maobin Xie, Chenlu Li, Jingyi Yu, Shaowei Wang* and Wei Lu. “Short-Wave Infrared Chip-Spectrometer by Using Laser Direct-Writing Grayscale Lithography.” *Advanced Optical Materials(AOM)*. 2022.
- (6) Guoliang Tang, **Zi Wang**, Shijie Liu, Chunlai Li* and Jianyu Wang*. “Real-Time Hyperspectral Video Acquisition with Coded Slits.” *Sensors*. 2022.
- (7) Yanhua Yu, Siyuan Shen, **Zi Wang**, Binbin Huang, Yuehan Wang, Xingyue Peng, Suan Xia, Ping Liu, Ruiqian Li and Shiyong Li*. “Enhancing Non-line-of-sight Imaging via Learnable Inverse Kernel and Attention Mechanisms.” 2023 IEEE/CVF International Conference on Computer Vision (ICCV). 2023.
- (8) Qingquan Liu, Zhiyi Xuan, **Zi Wang**, Xinchao Zhao, Zhiqin Yin, Chenlu Li, Gang Chen, Shaowei Wang*, and Wei Lu. Low-cost micro-spectrometer based

on a nano-imprint and spectral-feature reconstruction algorithm. *Optics Letters*, 2022, 47(11): 2923-2926.

- (9) Zhengqing Pan, Ruiqian Li, Tian Gao, **Zi Wang**, Siyuan Shen, Ping Liu, Tao Wu, Jingyi Yu and Shiyong Li*. "Onsite Non-Line-of-Sight Imaging via Online Calibration." *IEEE Photonics Journal*. 2022.

申请或已获得的专利:

Siyuan Shen, **Zi Wang**, Shiyong Li, Jingyi Yu. "Non-line-of-sight imaging via neural transient field." U.S. Patent Application No. 18/574,049.

Zi Wang, Wei Lu, Jingyi Yu. "Neural Spectral Fields Reconstruction for Spectrometer." Worldwide Patent Applications, PCT/CN2021/105312.

参加的研究项目及获奖情况:

上海市科学仪器专项-用于危险气体分子光谱多要素探测的热红外高光谱视频相机。(参与)

基于计算的高空间-光谱分辨率成像技术。(参与)